FACULTY OF **PHARMACEUTICAL SCIENCES**

HISTONE CODE CAUTION: CLIPPING AND PROPER PROPIONYLATION

Apr. Paulien Meert

Laboratory for Pharmaceutical Biotechnology
Faculty of Pharmaceutical Sciences
Ghent University

Promoters: Prof. dr. apr. Dieter Deforce
Dr. Maarten Dhaenens
Dr. apr. Liesbeth Vossaert

Thesis submitted in the fulfillment of the requirements for the degree of Doctor in Pharmaceutical Sciences
2016

MEMBERS OF THE EXAMINATION COMMITTEE

Prof. dr. Claudia Spits
Department of Embryology and Genetics, Faculty of Medicine and Pharmacy, Free University of Brussels

Prof. dr. Lennart Martens
VIB Medical Biotechnology Center, Ghent University

Prof. dr. Bart Devreese
Laboratory for Protein Biochemistry and Biomolecular Engineering, Faculty of Sciences, Ghent University

Prof. dr. apr. Sarah De Saeger
Laboratory of food analysis, Faculty of Pharmaceutical Sciences, Ghent University

Prof. dr. apr. Bart De Spiegeleer
Laboratory of Drug Quality and Registration, Faculty of Pharmaceutical Sciences, Ghent University

This research was funded by a Ph.D. grant of the intitute for the promotion of innovation through science and technology in Flanders (IWT Vlaanderen)

ACKNOWLEDGEMENTS – DANKWOORD

Hét dankwoord. Het is zover. Eindelijk mag ik iedereen bedanken die heeft bijgedragen aan dit boekje. En dat zijn er heel wat! Een doctoraat is namelijk, in tegenstelling tot wat de grammatica doet vermoeden, een werkwoord. Een werkwoord waar heel wat helpende handen bij nodig zijn. Al deze handen zou ik dan ook graag willen bedanken. Een mens zou niet verwachten dat dit bescheiden onderdeel van een thesis het moeilijkste neer te pennen valt. Toch is niets minder waar. Ik heb vermoedelijk zo een 5-tal versies op mijn computer staan, maar zelfs nu blijf ik stilletjes hopen dat ik niemand ten onrechte vergeten ben. Als dit wel zo is, denk er gerust enkele complimentjes bij!

Ook als het op bedankingen neerkomt, wil ik graag beginnen bij het begin. En dat is zonder twijfel bij mijn promotor. Dieter, ik wil je graag bedanken voor het vertrouwen dat je me hebt gegeven om hier, op lab FBT, te starten. Je hebt me steeds de vrijheid gelaten om allerlei hypothesen achterna te gaan, maar toch bijgestuurd waar nodig. Het leidt geen twijfel dat er zonder jou nooit sprake geweest zou zijn van dit boekje. Je had echter niet alleen oog voor het wetenschappelijke, maar hebt op ons op tijd en stond ook eens getrakteerd op een uitstapje, BBQ, ski-sessie, fietstocht, … Ideaal om de geest vrij te maken. Bedankt!

Van de ene promotor is het maar een kleine sprong naar de volgende. Maarten, jij hebt me tijdens mijn thesisperiode op lab FBT al meteen geïnspireerd met je eindeloze enthousiasme en liefde voor de wetenschap. Het siert je dat een MS spectrum je minstens even gelukkig kan maken als de eerste zonnestralen in april, die ons naar de bankjes voor het FFW loodsen over de middag. Daar vragen we dan beide "alle groentjes" op ons slaatje/broodje, ook al weten we maar al te goed dat er ons een "zoldersmaak" te wachten staat. Het is een kwestie van volharding, en dat heb ik zeker en vast van jou geleerd!

En dan de derde promotor. Dat is een "specialleke". Liesbeth, jij bent voor mij een promotor geweest in alle betekenissen van het woord. Je hebt me van bij de start met hand en tand begeleid en geleerd hoe ik de daad bij het woord moest voegen. Ondanks het feit dat je nu de officiële titel "promotor" draagt, blijf je in mijn hoofd "gewoon" een vriendin voor het leven. Zo ééntje waar je hoogtepunten en dieptepunten mee beleefd hebt. Zo ééntje waar je altijd bij terecht kan, vooral na de werkuren. Zo ééntje waar je gewoon ronduit je gedacht tegen mag zeggen (en dat kunnen we nu net allebei zo goed!). Zo ééntje waarmee je steeds "de grootste badkamer van het hotel" mag delen. Zo ééntje waar je enorm dankbaar voor bent!

Verder wil ik ook mijn dank betuigen aan de andere leden van de examencommissie. Hartelijk dank om mijn thesis zo uitvoerig te evalueren en een constructieve bijdrage te leveren.

Als ik terugdenk aan de voorbije 5 jaar, dan ben ik ook enorm dankbaar dat ik deel heb mogen uitmaken van zo een enthousiast en gedreven team. Als andere mensen al eens hun beklag deden over hun collega's, dan kon ik niet anders dan zeggen: "Heb ik geen last van.". Wat een luxe!

Eerst en vooral, mijn "compagnon de route", Bart. Samen gestart op een zonnige dag in augustus, samen geklonken op onze IWT beurs, samen gestrest tussendoor en nu (quasi) samen afgelegd. Het hoeft geen betoog dat samen altijd leuker is dan alleen! En al zeker met iemand die altijd goed gezind is, steeds bereid te helpen en daarbovenop ook nog eens over de juiste kwaliteiten beschikt om voor "den druivelaar" te gaan schrijven. Merci!

Daarnaast zijn er nog heel wat andere mensen waar ik - van ver of dicht - mee heb mogen samenwerken. Ellen, bedankt voor de vele fijne uren in het L-labo bij de start van ons doctoraat, waarna we ons beide meer richting massakot verplaatsten. De babbels bleven echter altijd even ongedwongen! In dat massakot kwamen we Sofie vaak tegen, de zoveelste naald aan het vervangen of aan het tunen. Sofie, bedankt om steeds zo ongedwongen jezelf te zijn. Rechttoe, rechtaan, dat is helemaal mijn stijl! Ook Franck was vaak achter de hoek van het massakot te vinden. Merci om zo zorgzaam met "onze kindjes" om te springen en 'chapeau' voor je onophoudelijke inspanning om Nederlands te spreken. Elisabeth, een andere bewoner van het massakot, ook al ben je ondertussen al aan je derde jaar bezig, je brengt nog altijd een frisse wind door het labo. Het klikte meteen tussen ons, en ik ben er dan ook van overtuigd dat het geen toeval is dat we beide zo graag naar "den tien minuutjes" gaan. Laura en Sander, jullie zijn de nieuwste generatie "histon people": doe dat goed en niet vergeten dat het nu eenmaal de meest ingewikkelde eiwitten ooit zijn!

In één adem beland ik zo bij den (ex-)bokaal, de mensen die - ondanks de noise cancelling headphones – alle pieken en dalen eigen aan het academisch onderzoek van dichtbij meemaakten: Katleen (een eindeloze bron van goede raad), Trees (organisatorisch talent en mede groene-mapjes liefhebber), Marlies (no nonsense en met twee voeten op de grond), Sandra (steeds vol interesse), Veerle (biedt een fris perspectief), Pieter (kan volledig in iets opgaan, om het daarna ten gepaste tijde te relativeren). Bedankt voor de vele leuke babbels over de middag en daarbuiten. Jullie brachten op tijd en stond "de luchtigheid" die nodig is om te volharden. Thesistudenten, jullie hebben niet alleen het spreekwoord "vele handen maken licht werk" in de praktijk omgezet, jullie boden ook steeds een frisse kijk. Bedankt Stefi, Leen, Maarten, Ine, Sarah, Sandra en Sebastiaan.

Maar lab FBT bevindt zich natuurlijk niet alleen op het tweede verdiep. Ook op het derde zit er een mooie afvaardiging. Lieselot en Senne, jullie zijn beide sfeermakers en gedreven in jullie werk. Bedankt! Op het gelijkvloers zit "den IT". Steeds paraat om netwerkproblemen op de lossen en menige harde

schijf te overbelasten. Behalve dan op woensdag 12 uur, dan kan je ervan op aan dat de spaghetti van "den Boudewijn" eraan moet geloven. Filip, Dieter, Christophe, Yannick en Bert, merci om op alle andere weekdagen de middagpauzes mee op te fleuren. Hoewel, ik ben te voorbarig als ik zeg dat alleen "den IT" op het gelijkvloers zit. Mado, jij hebt als enige vrouw enkele jaren je mannetje gestaan tussen al dat testosteron en deed dat steeds met glans. Je bruist van ambitie, en daar bewonder ik je voor!

Naast het research-gedeelte wil ik ook graag de mensen van het DNA-labo bedanken. Astrid, Nadine en Inge, bedankt om de administratieve zaken op jullie te nemen. Saskia, Petra, Sabine, Sylvie, Eveline, Evelien en Leen, bedankt voor de leuke babbels in de wandelgangen. David, mijn loop-buddy, merci om me bijna zelf te doen geloven dat "je nu eenmaal zelf ook niet sneller wil lopen". Zowel ik als je loop-app geloven er niets van! Ook bedankt om me steeds scherp te houden met een vinnige opmerking. Ik doe niets liever dan een nog vinniger antwoord bedenken!

Werk en ontspanning moeten hand in hand gaan. Voor dat laatste kon ik terugvallen op "the girlies". We kennen mekaar al sinds het middelbaar en nog steeds slagen we erin even zot te doen als toen! Sandy, merci om steeds een luisterend oor en bron van goede raad te zijn. Knolie, altijd begripvol en enthousiast. Net wat ik soms nodig heb! Katie, een echte spring-in-'t-veld die altijd lacht. Merci om alles af en toe eens wat "lichter te maken".

Als ik daarnet schreef dat ik zou "beginnen bij het begin", dan was dat misschien wat te voorbarig. In de letterlijke zin start alles namelijk bij mijn twee ouders. Mama en va, bedankt om me te laten opgroeien in een warm nest waar ik steeds de vrijheid heb gekregen om te doen wat ik graag doe. En neen va, zelfs als doctor kan ik je nog steeds niet tot op moleculair niveau verklaren hoe dat vuur in onze stoof zich gaande houdt. Er moeten nog doelen zijn in het leven he. En dan is er natuurlijk nog die andere sfeerbrenger in huis: "onze Jef". Broer, merci om de typische kuren van een grote zus steeds met de mantel der liefde ter overdekken en altijd je positieve zelf te blijven! Eline, als jij over literatuur begint, dan besef ik telkens dat ik nog zoveel te leren heb. Indrukwekkend. Jullie zijn een mooi koppel! Meter, bompa, Marc, peter Paul, meter Jeanine, tantes en nonkels, neven en nichten, ook jullie wil ik graag bedanken voor de vele leuke feestjes, waar geregeld de geluidsnorm overschreden wordt. Ik vind het geweldig dat we er zo goed in slagen de familiebanden aan te houden.

En dan is er natuurlijk nog "de KWP". Al sinds 2003 ben ik bij jullie kind aan huis. Eerst via de voordeur, maar al gauw werd ik gepromoveerd tot achterdeur-gebruiker. Marc en Annelies, bedankt om er steeds te staan voor Silas en mij en al die kaarsjes te doen branden wanneer nodig. Hymne, een unieke naam voor een unieke schoonzus! Thijs, bedankt om te supporteren vanuit Dubai. Loïc en Sarah, wij kijken vol spanning mee uit naar de geboorte van jullie eerste kindje. Jullie gaan dat super doen! Familie Rydant

en familie Zwaenepoel, ook bij jullie kom ik al 13 jaar over de vloer en voel ik me als een vis in het water. Bedankt daarvoor!

Ik verschiet ervan hoeveel pagina's ik hier al vol "bedankt" heb, wetende dat de hoofdrolspeler eigenlijk nog op scène moet verschijnen. Silas, mijn schatsietje, jouw naam is al terug te vinden op mijn ringmappen van het middelbaar (consequent met een hartje boven de i). Het is eigenlijk bijna niet te vatten dat je al bijna de helft van mijn leven naast mijn zijde staat. Ik weet duidelijk wanneer ik "een goeke" te pakken heb. Bedankt voor de kleine schouderklopjes, de pragmatische kijk en het luisterend oor. Het voelt wat oppervlakkig om iemand te bedanken voor alles, maar in jouw geval is dat anders: bedankt voor ALLES!

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

5-caC	5-carboxylcytosine (5-caC)
5-fC	5-formylcytosine (5-fC)
5-hmC	5-hydroxymethylcytosine
5-mC	5-hydroxymethylcytosine
A	Alanine
Aa	Amino acid
Ab	Antibody
ABC	Ammonium bicarbonate
Ac	Acetylation
ACN	Acetonitril
AEBSF	4-(2-Aminoethyl)benzenesulfonyl fluoride hydrochloride
ANOVA	Analysis of variance
AQUA	Absolute protein quantitation
AUC	Area under the curve
B2M	*β-2-microglobulin*
bFGF / FGF-2	Basic fibroblast growth factor
BH	Bovine histones
BMP	Bone morphogenetic protein
BSA	Bovine serum albumin
C	Cysteine
cDNA	Complementary DNA
CENP-A	Histone H3-like centromeric protein A
cH3	Clipped histone H3
CID	Collision induced dissociation
CpG	Cytosine-phosphate-Guanine
D	Aspartic acid
Da	Dalton
DDA	Data dependent acquisition
DIA	Data independent acquisition
DMEM	Dulbecco's modified Eagle medium
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DP	Desired product
E	Glutamic acid
E8	Essential 8
EB	Embryonic body
ECD	Electron-capture dissociation
EDTA	Ethylenediaminetetraacetic acid
eGFP	Enhanced green fluorescent protein
ESC	Embryonic stem cells
ESI	Electrospray ionization
ETD	Electron-transfer dissociation

FDR	False discovery rate
FMDV	Foot-and-Mouth disease virus
G	Glycine
H	Histidine
H3.cs1 Ab	Histone H3 cleavage site 1 Ab (antigen alanine 21)
H3K27me3	Histone H3 lysine 27 trimethylation
H3K4me3	Histone H3 lysine 4 trimethylation
HA	Hydroxylamine
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
hESC	Human embryonic stem cells
hmSILAC	heavy methyl SILAC
hPTM	Posttranslational histone modification
HRP	Horseradish peroxidase
I	Isoleucine
ICAT	Isotope-coded affinity tag
ICM	Inner cell mass
ID	Identification
IPA	Isopropylalcohol
iPSC	Induced Pluripotent stem cells
ISPTM	Iterative search for identifying PTMs
iTRAQ	Isobaric tags for relative and absolute quantification
K	Lysine
KLF-4	Kruppel-like factor 4
Kpna2	Karyopherin alpha 2
L	Leucine
LC	Liquid chromatography
LCMS	Liquid chromatography coupled to mass spectrometry
LIF	Leukemia inhibitory factor
lncRNA	Long non-coding RNA
M	Methionine
m/z	Mass to charge ratio
MALDI	Matrix assisted laser desorption ionisation
Mcm4	Minichromosome maintenance complex component 4
MD-score	Mascot delta score
Me	Methylation
MEFs	Mouse embryonic fibroblasts
mEpiSC	Mouse epiblast stem cells
mESC	Mouse embryonic stem cells
miRNA	Micro RNA
mRNA	Messenger RNA
MS	Mass spectrometry
MSMS	Tandem MS
MW	Molecular weight
N	Asparagine

NCBI	National Center for Biotechnology Information
ncRNA	Non-coding RNA
NE	Nuclear extract
Oct4	Octamer-binding transcription factor 4
OP	Overpropionylated product
P	Phenylalanine
PA-reagent	Propionic anhydride reagent
PBS	Phosphate buffered saline
PCA	Principal component analysis
PCR	Polymerase chain reaction
Ph	Phosphorylation
PI	Propidium iodide
PIC	Protease inhibitor cocktail
PICMI	Post-inner cell mass intermediate
piRNA	Piwi-interacting RNA
POU5F1	*POU class 5 homeobox 1*
Prop. Anh	Propionic anhydride
PTM	Posttranslational modification
PVDF	Polyvinylidene fluoride
Q	Glutamine
Q-TOF	Quadrupole – Time-of-flight
R	Arginine
RA	Retinoic acid
RNA	Ribonucleic acid
RP-HPLC	Reversed phase high pressure liquid chromatography
RPL13A	*Ribosomal protein L13A*
RP-LC	Reversed phase liquid chromatography
RPMI	Roswell Park Memorial Institute
RT-qPCR	Reverse transcription – quantitative PCR
S	Serine
Sall4	Spalt-like transcription factor 4
SAM	S-adenosyl methionine
SDS	Sodium dodecyl sulphate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
SILAC	Stable isotope labeling by amino acids in cell culture
siRNA	Short interfering RNA
Sox2	(Sex determining region Y)-Box 2
SSEA1, 3, 4	Stage-specific embryonic antigen 1, 3, 4
SWATH	Sequential Window Acquisition of all THeoretical Mass Spectra
T	Threonine
Taq	*Thermus aquaticus*
TEAB	Triethylammoniumbicarbonate
TEB	Triton-X extraction buffer
TERT	Telomerase reverse transcriptase

TET	Ten-11 translocation
TMT	Tandem mass tags
Top2A	Topoisomerase (DNA) II alpha
Tra	Keratan sulfate antigen
Tris	Tris(hydroxymethyl)aminomethane
UP	Underpropionylated product
V	Valine
W	Tryptohan
XIC	Extracted ion chromatogram
Y	Tyrosine

"You start out as a single cell derived from the coupling of a sperm and an egg; this divides in two, then four, then eight, and so on … The mere existence of such a cell should be one of the great astonishments of the earth. People ought to be walking around all day, all through their waking hours calling to each other in endless wonderment, talking of nothing except that cell."

— Lewis Thomas (1979), the Medusa and the snail

CHAPTER 1:

BIOLOGICAL INTRODUCTION

1. BIOLOGICAL INTRODUCTION

1.1. Human embryonic stem cells

1.1.1. A definition

Human embryonic stem cells (hESC) are distinguished from other cell types by two main properties. First, their prolonged proliferation potential: hESC can be cultured *in vitro* and expanded in number indefinitely while maintaining the undifferentiated state characteristic of the blastocyst's cells from which they are derived. Second, both *in vivo* and *in vitro* they have the remarkable capacity to develop into all different cell types of the human body.

1.1.2. Derivation of an hESC line and appropriate culture conditions

Human ESC lines are derived from the blastocyst's inner cell mass (ICM), generated by *in vitro* fertilization. Such embryos that are left unused can be donated for research purposes. Once the blastocyst stage is reached five days after fertilization, cells of the ICM are transferred from the pre-implantation stage embryo into a plastic laboratory culture dish. In the original protocol of J. Thomson, the culture dish was coated with a feeder layer of mouse embryonic fibroblasts (MEFs) (Figure 1.1) [1]. Nevertheless, nowadays hESC can also be cultured in feeder-free conditions [2].

In case MEFs are used as feeder layer, they are first mitotically inactivated with e.g. mitomycin C, in order not to overgrow the ESC. This is a potent chemotherapeutic agent that acts through DNA crosslinking, to inhibit DNA synthesis and cell division. These feeder cells are used because they support ESC proliferation and prevent spontaneous differentiation. The latter is accomplished by secretion of growth factors such as basic fibroblast growth factor (bFGF or FGF-2), transforming growth factor β, activin A, extracellular matrix proteins and antagonists of the bone morphogenetic protein (BMP) signaling pathway [3].

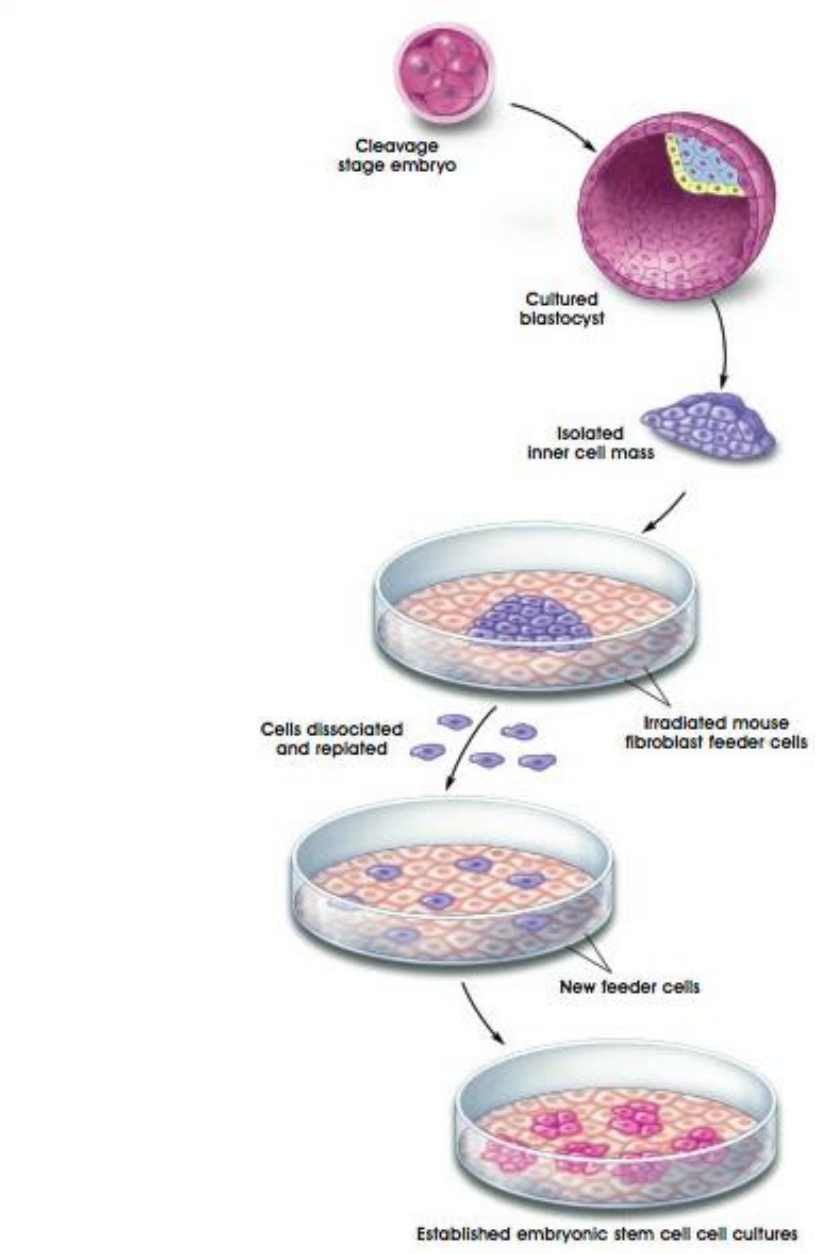


Figure 1.1 Establishing a hESC culture on a MEF feeder layer [4]
To generate hESC, cells derived from the blastocyst's inner ICM are cultured in a multi-step process.

1.1.3. **ESC differentiation: from totipotent to unipotent**

Stem cells can differentiate into any of the hundreds of cell types in the human body. Yet, different potency levels are found throughout this differentiation process, whereby cells become more and more specialized at each step (Figure 1.2). The singe-cell zygote is totipotent and thus capable of forming both embryonic and extra-embryonic tissues. Five days after fertilization the human blastocyst is formed. The ICM and the ESC derived thereof are pluripotent. These cells can still differentiate into germ cells and any of the three germ layers: endoderm (lungs, liver and pancreas) , exoderm (nervous system and epidermal tissue) and mesoderm (heart, blood, muscle and bone) [5][6]. These three germ layers are the

embryonic source of all cells of the body. Upon further differentiation, cells become multipotent and can develop only in a limited range of cells within a tissue type, thereby differentiating towards oligopotent cells. These so-called progenitor cells can give rise to only a few cell types of a particular lineage. For example, a lymphoid stem cell is an example thereof: it can give rise to blood cells of the lymphatic system (T, B and NK cells), but can no longer develop into any kind of blood cell like the multipotent bone marrow stem cells can. The least potent precursor cells are called unipotent. These cells still have unlimited reproductive capacities but can only give rise to one specific cell type. For example, most epithelial tissues can renew damaged cells throughout adult life due to the presence of unipotent stem cells [7].

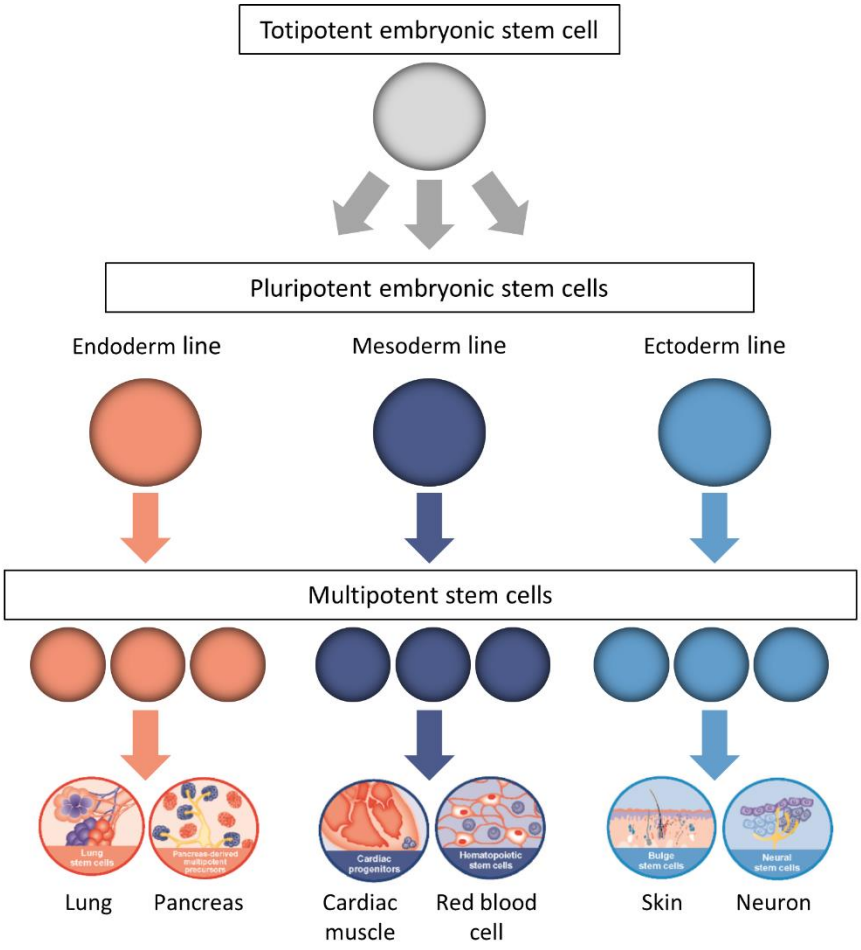


Figure 1.2 Differentiation of an embryonic stem cell

1.1.4. **Characterization of hESC**

Numerous hESC lines have been derived worldwide since the first derivatization in 1998 by Thomson [1]. Since all these hESC lines are genetically different and moreover cultured in diverse conditions, it's not surprising that they are mutually divergent in morphology, growth dynamics and tendency to

differentiate spontaneously. Nevertheless, it remains crucial to be able to define the characteristics of a "true" hESC. This profiling can be done using 4 different types of markers: extracellular cell-surface markers, intracellular markers (e.g. transcription factors), general protein profile and epigenetic features [8].

Extracellular markers have proven to be a powerful tool for distinguishing undifferentiated stem cells from their differentiated counterpart. It is interesting to know that some markers such as SSEA-1 are expressed only in mouse ESC, while other markers such as SSEA-3 and SSEA-4 are specific for hESC [9]. Besides these extracellular markers, monitoring the gene expression of specific markers is often used as an intracellular marker for hESC, thereby highlighting pathways unique for hESC. The expression pattern of genes coding for the transcription factors Nanog and Oct4 are typical markers of pluripotency [10]. A third way to define pluripotency is by the protein content. Van Hoof et al. [11] compared the proteomes of undifferentiated hESCs with their derivatives formed after 12 days of undirected "spontaneous" differentiation into a heterogeneous population. More than 700 proteins were identified as being present in undifferentiated cells only. Among these were several proteins that at the time were not previously known to be enriched in or specific for ESCs, such as: Top2A, Mcm4, Kpna2, and Sall4. Although both mass spectrometry (MS) based proteomics and stem cell research are upcoming fields, it is still difficult to obtain sufficient tissue to combine genomic and quantitative proteomics data. A final way of defining stemness is by using epigenetic markers, such as the chromatin state, which will be discussed in more detail in section 1.5.2.

1.2. Induced pluripotent stem cells: maximize your pluripotential

Although cell fates during development are neither restrictive nor irreversible, many assumed terminally differentiated cells had lost the potential of producing other cell types. Yamanaka and co-workers proved otherwise when they established the first induced pluripotent stem cells (iPSC) from mouse fibroblasts [12] in 2006. These iPSC are pluripotent cells generated from adult somatic cells by means of transfecting only four transcriptional factors: *OCT4, SOX2, c-MYC* and *KLF4*. Just six years later, this groundbreaking discovery would get him to win the Nobel Prize in Physiology or Medicine. In 2007, the same group published a study where iPSC were derived that could generate viable chimeras when injected in embryos, in contrast to the first generation iPSC. This is important since the formation of viable chimeras is considered a crucial characteristic for pluripotent stem cells. Therefore, they had to use Nanog as a marker to select for pluripotent stem cells instead of Fbx 15. At the same time, Thomson et al. achieved to create human iPSC out of human fibroblasts by means of a slightly different cocktail: *OCT4, SOX2, c-MYC* and *LIN28* [13].

Despite the enormous potential for future applications in regenerative medicine, safety is the major issue burdening the clinical implementation of iPSC. Conventional reprogramming strategies rely on the stable integration of transgenes but coordinately introduce the risk of mutations being inserted into the target cell's genome [14]. Hence, several nonintegrating reprogramming techniques have been developed to circumvent the risk of spontaneous tumor formation and to improve the quality of the generated iPSC, including the almost complete removal of the integrated viral DNA or alternatively, the use of nonintegrating viruses [15,16]. Furthermore, genome engineering is expected to further improve iPSC quality [17]. These virus-independent reprogramming methods are based on DNA, protein, or messenger RNA (mRNA) expression [18–20].

1.3. Stem cell states: PICMI as the connecting link between naïve and primed pluripotency

Embryonic cells are pluripotent for only a short window of time in the embryo, nevertheless it has been shown that pluripotency is not a fixed state. Two phases of pluripotency can be defined: naïve – in the blastocyst and primed – in the post implantation epiblast [21–23].

In theory, hESC should share more features with mouse ESCs (mESC), since both these cells are derived from similar stages in embryo development. Instead, hESCs - as well as human iPSC - appear to share defining features with mouse epiblast stem cells (mEpiSCs), such as a flattened morphology and epigenetic silencing of one of the two female X chromosmes, amongst other things (Table 1.1). Nonetheless, both pluripotent states share a common gene expression signature which encompasses the core pluripotency network including *Oct4*, *Sox2* and *Nanog*. Of relevance, detailed analysis of the morphological changes of the human ICM during the hESC derivation process revealed the formation of an obligatory transient structure, coined as the post-inner cell mass intermediate or PICMI, which expresses both early and late epiblast cell markers [24–26]. Further investigations on the characteristics of the PICMI, will help to get a more clear understanding of the human pluripotent states.

Table 1.1 Comparison of pluripotent stem cell types based on commonly recognized cell properties [27]

Properties	**mESC NAIVE**	**mEpiSC PRIMED**	**hESC PRIMED**
Origin	ICM	Epiblast	ICM
Morphology	Domed	Flattened	Flattened
Clonogenicity	High efficiecy	Low efficiency	Low efficiency
Single cell clonogenicity	Yes	No	No
Passaging	Insensitive	Sensitive	Sensitive
Chimerism	High efficiency	Very low efficiency	N/A
Growth factor dependence	LIF and BMP4	bFGF and Activin	bFGF and Activin
Female X inactivation	XaXa	XaXi	XaXi
Markers	*Oct4, Sox2, Nanog, Ssea-1, Rex-1, Fgf4, Tert, klf2, Klf4, Stella*, …	*Oct4, Sox2, Nanog, Ssea-1, Fgf5, …*	*OCT4, SOX2, NANOG, SSEA-3/4, TRA-1-60, TRA-1-81, REX-1, TERT, STELLA, …*

1.4. The ethical dilemma of human stem cell research

Few people doubt the medical potential of human stem cell research. Hence, the main controversy surrounding stem cell research is not *whether* we should use stem cells for research and therapeutic purposes, but rather about *what* sort of stem cells we should use and *how* we should obtain them.

As with all other newcomers in the medical field, stem cell research deals with issues such as proving the effectiveness, minimizing the risks, personal questions about donation and consent, and social issues like the price tag being too high for patients or health services. Nonetheless, for the majority of people, the most difficult dilemmas arise regarding the downstream effects of hESC research (Table 1.2). Either one supports this and accepts the embryo destruction that comes along with it, or either one opposes and thereby accepts potential benefits of this research will be foregone [28]. Many people however, fall somewhere in between these opposite points of view. They see the human need and hope for new treatments, but at the same time believe a human embryo is more than just a clump of cells.

Table 1.2 Ethical issues at different phases of stem cell research [29]

Phase of research	Ethical issues
Donation of biological materials	Informed and voluntary consent
Research with hESC	Destruction of embryos
	Creation of embryos specifically for research purposes
	Payment to oocyte donors
	Medical risks of oocyte retrieval
	Protecting reproductive interest of women in infertility treatment
Use of stem cell lines derived at another institution	Conflicting legal and ethical standards
Stem cell clinical trials	Risks and benefits of experimental intervention
	Informed consent

In order to form an opinion, one can ask himself: "What moral status does the human embryo have?". Some people believe that an embryo is a person with equal moral status as a baby or an adult. They feel we should not judge the embryo by its state of development but by what it will become. The opposite view says that the embryo becomes a person in the moral sense only at a later stage of development than fertilization. Some are even convinced that hESC research is a moral duty of humanity, if it provides a potential means to find treatments for otherwise serious and incurable suffering. But many other people believe neither of these radically opposed views do justice to the complexity of the developing embryo and hold a middle ground. They believe that the early embryo has a special status as potential human being but that it is acceptable to use it for research purposes given there is good scientific justification, careful oversight and informed consent from the couple donating the embryo for research [29,30].

Does the dilemma get any easier if we move on from the theoretical discussion to a concrete example? Even though the first steps in that direction are made, cell replacement therapies based on hESC have a long way to go. But what if the use of hESC could reduce the use of animals for testing? Human stem cells and their derivatives could provide virtually unlimited sources of tissue to test drugs and chemicals for toxicity [31]. That way, the use of animals can be limited and gained insights can lead to improved human prediction models. On the other hand, can we replace one ethically controversial process (using animals) by another (using hESC)? If one disapproves of using animals to test cosmetics, can the use of cells derived from human embryos be approved instead?

As the ample use of question marks used in this section illustrates, hESC research raises some complex moral and social issues. As a result, not only the public but also the funders of such research expect scientists to be able to show they are aware of, and understand, the ethical and social dimensions of

their work. The ethical issues thus need to be considered, along with the scientific challenges to ensure stem cell research is carried out in the appropriate manner.

1.5. Epigenetics

1.5.1. A definition

All cells in an animal originate from the same cell and therefore hold the same genetic material. Yet, e.g. a brain cell and a liver cell use this genetic information in a different way. This means there is a regulating mechanism on top of genetics, called epigenetics in which "epi" is the Greek prefix for "above". When looking up the word "epigenetics", one will find several definitions coming to the same conclusion: "An epigenetic trait is a stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence itself" [32,33].

The epigenome consists out of several epigenetic marks that tell genes to switch on or off, to speak loudly or whisper. It is through these epigenetic marks that extracellular and environmental factors like diet, stress and medication can make an imprint on our genes, that is passed on through cell division and from one generation to the next. The discovery that acquired characteristics can be inherited is a relatively new concept. During most of the 20th century we had a deal with biology: whatever choices we made causing obesity or hasten death, we could always rely on our genes. Our children would take a fresh start, the genetic slate would be wiped clean. Nowadays it is half time in the big "nature versus nurture" play-off and the answer lies somewhere in between. Our health, personality and other characteristics are based largely on the genes we inherited, but daily life habits and environment can result in epigenetic changes, altering gene expression. This also explains the differences between genetically identical twins. Some of the strongest evidence for transgenerational inheritance in humans comes from the survivors of the Dutch Hunger Winter (November 1944 – spring 1945) during which the population was trying to survive on only about 30% of the normal daily calorie intake. Monitoring of the pregnant women revealed that not only the birth weight of their own child was influenced by malnutrition, but the birth weight of their grand children as well. But make no mistake, epigenetics might be a new revolution, it's not evolution nor does it dispel the main ideas of Darwinism. An epigenetic response can be inherited through many generations, but once the environmental pressure is removed it will eventually fade and the DNA code will return to its original programming.

Waddington, who is given credit for coining the term "epigenetics", created a metaphore for how gene regulation modulates development: the epigenetic landscape, as displayed in Figure 1.3 [34]. In this metaphore a marble rolling down a hill toward the lowest point represents the process of cellular decision making during differentiation and development. While still of great value in the context of

normal development, the Waddington model falls short of accommodating recent breakthroughs in cell reprogramming. The breakthrough of iPSC and advances in direct cell fate conversion suggest that somatic and pluripotent cell fates can be interconverted without transiting through distinct hierarchies [35]. The term epigenetics and its different components will be further discussed in section 1.5.2.

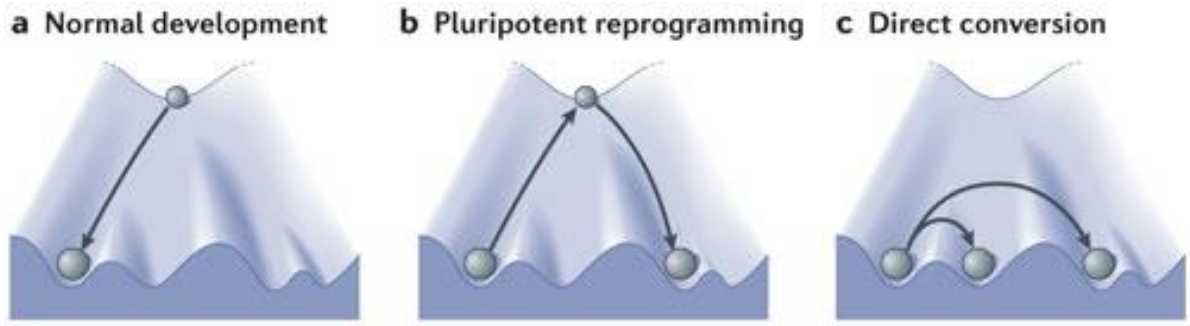


Figure 1.3 Waddington's epigenetic landscape [35]
A: Marbles rolling down a specific groove represent the lineage commitment of pluripotent stem cells during cell development. **B**: An iPSC is symbolized by a marble rolling from the bottom of the hill back to the top. Thereafter it can be redifferentiated in another specific somatic cell type. **C**: During transdifferentiation, cell fates can be directly interconverted.

1.5.2. **Unraveling the epigenetic layers**

The total length of the DNA in a single diploid human cell is approximately 2 m, but even so, it must fit into a cell's nucleus that is roughly 10 µm in diameter. To accomplish this, the structural organization of DNA into eukaryotic chromatin involves several orders of compaction. Nucleosomes are the basic DNA packaging elements, formed by DNA wrapped around small histone proteins [36]. Epigenetic effects can take place at the chromatin, the DNA as well as the histone level. Furthermore, there are also RNA-based epigenetic mechanisms. All these different epigenetic mechanisms are shown in Figure 1.4 and will be described in the following paragraphs.

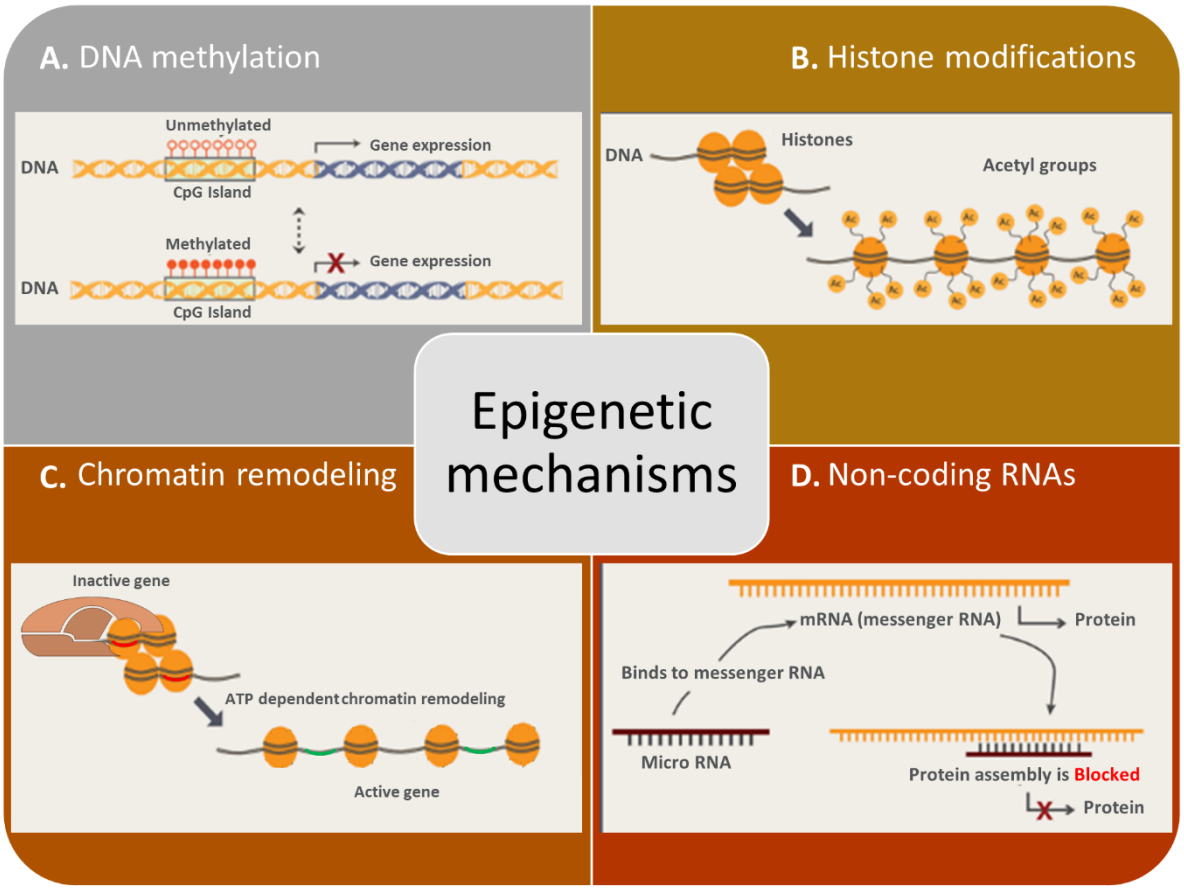


Figure 1.4 Epigenetic mechanisms
The different mechanisms involved in epigenetics can be divided into 4 main groups. **A:** DNA methylation will hamper gene expression. **B:** PTMs on a histone tail can both induce or repress transcription, depending on the type of modification and the amino residue it is placed on. **C:** ATP-dependent chromatin-remodeling complexes regulate transcription by either translocating, ejecting or restructuring nucleosomes. **D:** Many of the non-coding RNAs can interact with translation of mRNA into a protein.

1.5.2.1. **DNA methylation**

Methylation at the 5′ site of cytosine (5-mC) at CpG dinucleotides - also known as DNA methylation - is a major player in the process of epigenetic inheritance. DNA methylation of the mammalian genome is a repressive mark, which is directed and preserved by the actions of a DNA methyl transferase (DNMT) family – also called "writers" (Figure 1.5). The effect itself is established by recruitment of so-called "readers", the methyl-CpG-binding domain or by blocking the binding of transcriptional factors [37].

Until recently DNA methylation was thought of as stable and irreversible, as the removal of the methyl group of the 5′ site of cytosine is a thermodynamically unfavorable event. Nowadays it is known that DNA demethylation can occur, and this in two different ways: passively during DNA replication or actively through enzymatic activities. Passive DNA demethylation involves DNMTs being denied access to the

newly replicated DNA, which results in an unmethylated status. Active DNA demethylation on the other hand can be exerted via one out of the two currently known ways of operating. First, activation-induced cytidine deaminase has been shown to play a key role in genome-wide demethylation of mouse primordial germ cells [38]. Second, several research groups demonstrated the capacity of Ten-11 translocation (TET) family proteins to convert 5-mC to 5-hydroxymethylcytosine (5-hmC), 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC), resulting in base excision repair, in the absence of cell division [39,40].

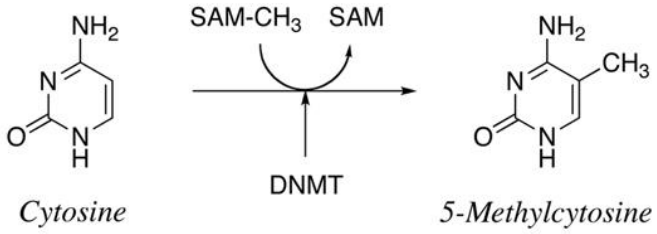


Figure 1.5 DNA methylation [41]
The DNMT family of enzymes catalyzes the transfer of a methyl group to DNA by using S-adenosyl methionine (SAM) as the methyl donor.

1.5.2.2. **RNA based epigenetic mechanisms**

Genome wide studies have revealed that even though only 1-2% of the human genome encodes for proteins, still up to 90% is transcribed [42]. Hence, this results in a large pool of non-coding RNAs (ncRNAs) which underlie an additional set of epigenetic mechanisms.

These ncRNAs can be classified into two main groups: infrastructural ncRNAs and regulatory ncRNAs. Ribosomal, transfer, small nuclear, and small nucleolar RNAs belong to the group of infrastructural ncRNAs. Regulatory ncRNAs can be divided into microRNAs (miRNAs), Piwi-interacting RNAs (piRNAs), small interfering RNAs (siRNAs) and long non-coding RNAs (lncRNAs) [43,44].

MiRNA are small RNAs (< 200 bp) that target mRNA in order to regulate gene expression. It has been postulated that they regulate the expression of 50% of all genes in a cell at posttranscriptional level [45]. In stem cells, it is shown that miRNAs are involved in controlling self-renewal and differentiation, by affecting DNA methylation processes [46]. SiRNAs can modulate gene silencing using the sequence complementary to the target mRNA, piRNAs on the other hand induce chromatin remodeling and transposon silencing by interacting with Piwi proteins. The majority of the ncRNAs belong to the group of lncRNAs, which comprise all ncRNAs of >200 bp in length. Among the most well described examples are the two lnRNAs Xist (17 kb) and its antisense transcript Tsix (40 kb) involved in X chromosome

inactivation, for which they ensure that only one of the two X chromosomes in females is expressed during development [43,44,46].

Familial traits, including physiological and developmental processes are not always transmitted according to the Mendelian rules. As part of the epigenetic machinery, the non-coding RNA might be another player involved in this process. In order to be heritable, RNAs must be present within ova or spermatozoa, or in both [47]. Several studies have shown that a complex and diverse set of RNAs is present within gametes of males and females, as well as in early embryos [48–52]. However, in order to correctly estimate the effect sizes of RNA heritability a lot more research is still needed.

1.5.2.3. **Chromatin based remodeling**

ATP-dependent chromatin remodeling complexes provide a means to package the same genetic material in two different chromatin states: euchromatin and heterochromatin. Euchromatin is the transcriptionally active state, whereas heterochromatin is the highly compacted form where only low levels of transcriptional activity are found. Together, the different subfamilies of chromatin-remodeling enzymes disrupt the chromatin interactions which allows for sliding the histone octamer across the DNA, changing the conformation of nucleosomal DNA and changing the composition of the histone octamer [53–55].

1.5.2.4. **Posttranslational histone modifications**

Since some posttranslational modifications (PTMs) on histones are inheritable during cell division, they are generally considered to be a major type of epigenetic marks [56]. In this section several topics concerning these histone PTMs will be discussed: composition of the nucleosome, main classification of histone PTMs, biological mechanism of histone PTMs and an overview of the main modifications.

As described above, in all eukaryotes, the genome is tightly associated with histone proteins in order to form chromatin, whose fundamental subunit is the nucleosome (Figure 1.6) [36]. Chromatin is essential for the compaction of genomic DNA but also represents a physical barrier to control DNA accessibility and gene expression. Each nucleosome is made up of four different types of core histones, which have been very well evolutionarily conserved [57]. Two H3:H4 dimers form a tetramer, which then binds to two H2A:H2B dimers in order to form an octamer. Due to the high abundance of lysine (K) and arginine (R), the alkaline histones are positively charged at physiological pH. DNA on the other hand has a negative charge coming from its phosphate group, therefore 146 bp of DNA can easily wrap around a histone octamer by mainly electrostatic forces. The nucleosomes, in turn, are held together by H1 linker histones [58].

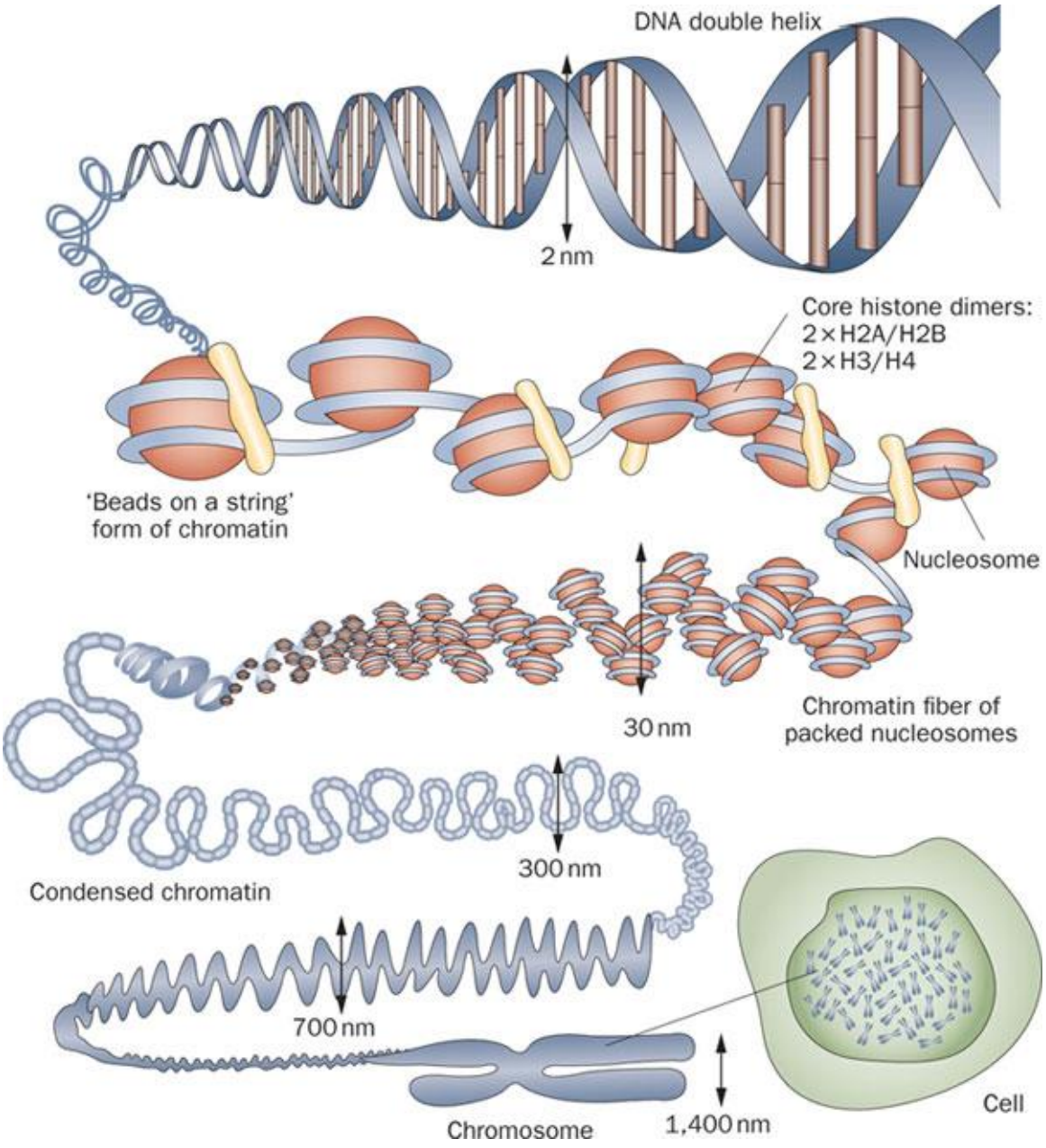


Figure 1.6 Condensed folding of DNA into a chromosome
The DNA double helix is wrapped around the core histone octamer (H2A, H2B, H3 and H4: two each) and forms a nucleosome. These fundamental subunits are packed together into chromatin. Chromatin is then further organized into a chromosome, located in the nucleus [59].

When looking at the structure of a nucleosome, one can see that the histone N-tails protrude out of the nucleosome, making them more susceptible for PTMs. Nevertheless, PTMs can also be found within the globular domain and even at the C-terminal tail of H2A and H2B [60]. Three different types of histone modifications can be distinguished. First, the addition of a functional group or protein to a specific amino acid (Aa). Examples thereof are K methylation, acetylation and ubiquitination amongst others. However, these PTMs represent only a small part of all the possible modifications, as probably some of them are yet to be discovered. Tan et al. published in 2011 a list of 67 new histone marks, which increased the

number of known histone modifications by 70% [61]. Second, the chemical or structural nature of the Aa residue itself can be altered by e.g. citrullination which is the posttranslational conversion (deimination) of an R residue within a protein to the non-coded Aa citrulline. An example of structural change is proline isomerization [62]. More drastic structural changes such as histone clipping are considered the third type of histone PTMs. The latter will be discussed more elaborately in Chapter 4. The combinatorial complexity of all different PTMs is often referred to as "the histone code" and extends the information potential of the genetic code [63,64].

Histone PTMs are thought to regulate chromatin structure and function by two main mechanisms. First, as a result of neutralizing the histones' charge state or through internucleosomal interactions, the higher-order structure of chromatin can be reformed, thereby influencing the access of DNA-binding proteins, such as transcription factors. Additionally, histone PTMs can function either by recruiting PTM-specific binding proteins ("readers") and their associated binding partners ("effectors") or by inhibiting the binding of a protein to the chromatin. Installing and removing of the histone PTMs themselves is regulated by a diverse group of enzymes, called "writers" and "erasers" respectively.

The PTM induced changes in interaction between chromatin and its binding factors is translated into a biological outcome [65]. Interestingly, the same histone PTMs can correlate with transcriptional activation or repression depending on the type and localization of the modification. Twenty types of histone PTMs have been reported (Figure 1.7) but only the main modifications will be discussed hereafter.

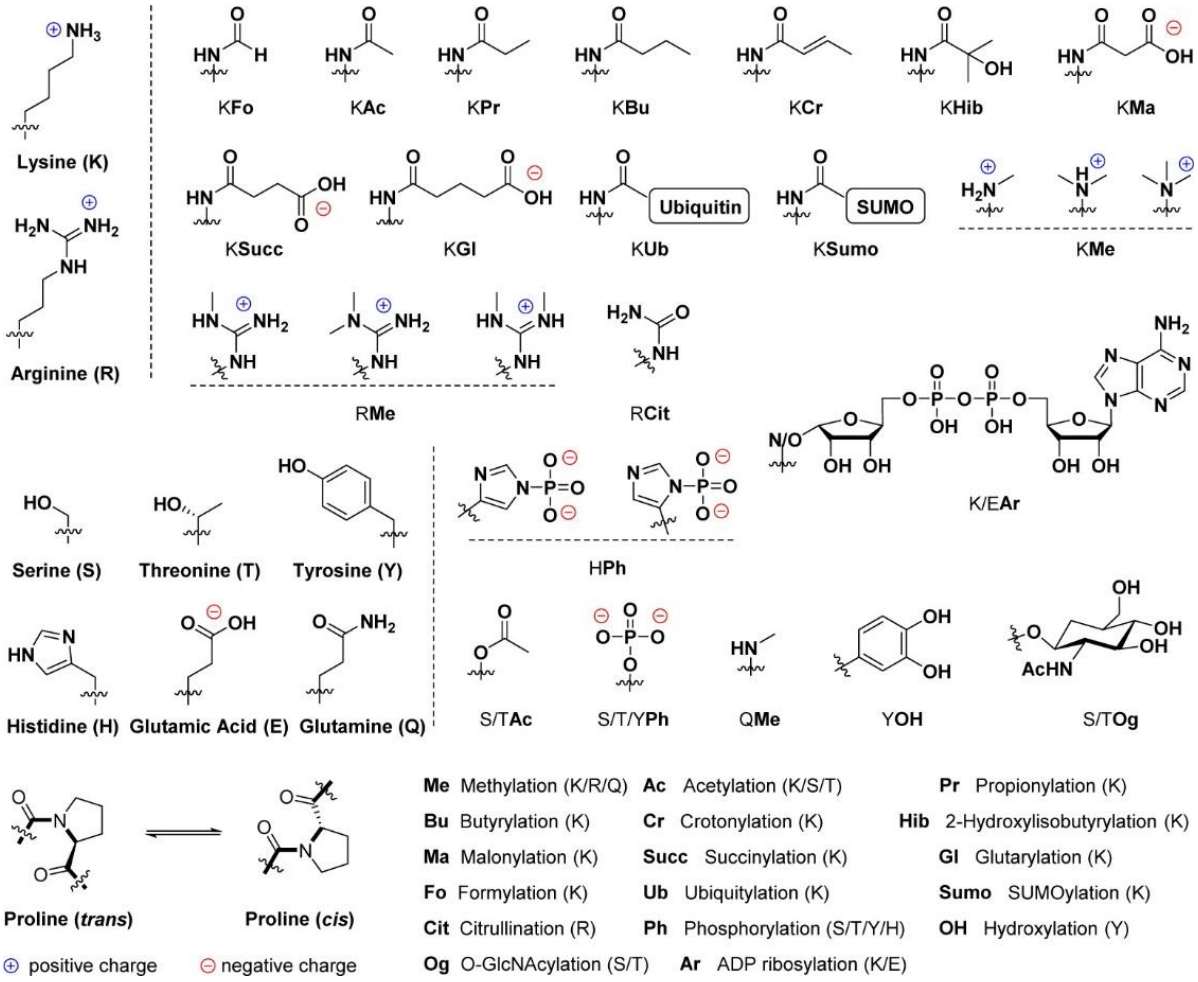


Figure 1.7 Structure of histone PTMs [66]

Acetylation (Ac) on K neutralizes the charge of the basic histone proteins and thereby interferes with the histone-DNA interaction. DNA becomes thus more accessible for DNA transcriptors, leading to increased gene expression. It is one of the most widely studied PTMs and is catalyzed by a group of enzymes called "histone acetyltransferases". Deacetylation on the other hand is catalyzed by "histone deacetylases" and results in gene silencing [67,68]. Acetylation increases the weight of a histone peptide by ~42 Da per acetyl moiety.

Lysine or arginine methylation (me) is a covalent modification that represents the addition of a methyl group from the donor SAM to the side-chain nitrogen residue. Although there is a pKa shift, methylation does not change the charge on histone proteins at physiological pH and hence its electrostatic interaction with DNA is not affected. It will mainly serve as an anchor site that can be recognized via specific binding domains. Thereby, methylation can be related to activation, elongation, or repression of gene expression dependent on the site and extent of methylation [67,69,70]. For example, H3K4me, me2 and me3 associate with transcriptional activation [71–73], whereas methylation on H3K27 leads to

transcriptional repression [74]. Histone arginine methylation is strongly associated with gene activation [67]. Methylation increases the weight of a histone peptide by ~14 Da.

Another important modification that occurs on histones is phosphorylation (ph) on serine (S), threonine (T) or tyrosine (Y) that is mediated by various specific protein kinases. Phosphatases on the other hand mediate the removal of a phosphate group. Just as acetylation, a phosphate group generates an overall negative charge on these residues, thereby interfering with DNA-histone interactions. Each phosphate group adds ~95 Da to the histone's molecular weight. Little is known about the relationship between histone phosphorylation and gene expression. Nevertheless, it is an important modification since histone phosphorylation participates in many cross talk events with other histone modifications. For instance, H3S28ph enables the demethylation and further acetylation of the neighbouring H3K27. Beside its role in chromosomal condensation and transcription, phosphorylation of H2A.X is a universal and instantaneous cellular response to the DNA damage [67,68,70].

Ubiquitination is a bulky modification (76 Aa protein molecule) on the small histone peptides. Each ubiquitin group adds approximately 8.5 kDa and in case of polyubiquitination multiple ubiquitin monomers are attached to a single lysine, thereby forming a polymer. Ubiquitination typically marks proteins for proteasomal degradation via the 26S proteasome but in case of histones it provides other signals important for gene regulation [67,69,70]. Three possible mechanisms were proposed for how histone ubiquitination affects transcription. First, due to its large size ubiquitin may affect chromatin folding and DNA accessibility. Second, the ubiquitin moiety may function as an anchor site for the recruitment of various regulatory molecules. Finaly, ubiquitination may also influence other histone modifications, thereby affecting transcription indirectly. For example, in euchromatin, ubiquitination would activate the transcription by enabling the methylation of H3K4 and facilitating the transcriptional elongation [70,75].

Like ubiquitylation, sumoylation is a very large modification (~100 Aa). It was demonstrated that sumoylation can take place on all four core histones. Sumoylation can also compete with other lysine-targeted modifications, including acetylation or ubiquitylation, and thereby can switch transcription from the active to the repressed state [68,70,76,77].

ADP-ribosylation is defined by the addition of poly (ADP-ribose) units specifically at the lysine residues. It plays an important role in DNA break repair mechanisms but also influences transcription. The latter is probably established through crosstalk with other histone modifications. For example, mono-ADP-ribosylation on H4 seems to occur preferentially when H4 is hyperacetylated [67,68,70].

1.5.2.5. **Histone variants**

Next to the four canonical histone proteins, many histone variant forms exist in different organisms. Histone variants differ in certain residues or regions of their Aa sequence compared to their canonical counterparts (Table 1.3) [78,79]. These variants can be expressed throughout the cell cycle, which suggests they have other functions to fulfill next to DNA compaction and are usually present as single copy genes. Since there are so many different variants and therefore combinatorial possibilities the question arises whether these structural variations induce differences in function and localization of the nucleosomal structure [80]. In general, H2B variants have tissue-specific functions, for example in the testes of both vertebrates and invertebrates [81–83]. The first H4 isoform has recently been identified in human fat cells [79]. H3 en H2A variants have a broad range of possible functions during development and differentiation, and some variants can switch places with pre-existing histones. Several groups reported the role of histone H3.3 in transcriptional activation [84,85].

Table 1.3 Overview of known histone variants [55,78–82]**.**

Canonical histone	Histone variants
H1	H1.0 H1.1 H1.2 H1.3 H1.4 H1.5 H1X
H2A	H2A.X H2A.Z H2A-BDB MacroH2A
H2B	hTSH2B H2BFWT TH2B
H3	H3.1 H3.2 H3.3 H3.t CENP-A
H4	H4.B

"Men are only so good as their technical developments allow them to be."

— George Orwell

CHAPTER 2:

TECHNICAL BACKGROUND

2. TECHNICAL BACKGROUND

2.1. The importance of proteomics

Proteomics is an emerging field in molecular biology (Figure 2.1) that deals with the systematic large scale analysis of proteins. The proteome is referred to as "all proteins expressed in a given organism, tissue, organ or cell culture under defined conditions".

Despite the great technological breakthroughs in the genomics field, such as next generation sequencing in 2005, the need for proteomics still remains. Nucleic acids, while undoubtedly vital molecules in a cell, are mainly information carriers. Therefore they can only tell us about protein function indirectly, while proteins are the actual functional molecules in the cell. Moreover, as RNA can have alternative splicing products and proteins can be modified in many different ways, over a million different proteoforms exist, in contrast to the ~20.300 protein-coding genes identified. Thus, much of the complexity created by our biological machinery originates at the level of protein variation rather than genome variation [86,87].

Proteins are omnipresent in all organisms as hormones, enzymes, antibodies, transporters, receptors, etc. and are thus involved in almost every biological function. A comprehensive analysis of the proteins in a cell provides a unique global perspective on how these molecules interact and cooperate. In addition, proteins are the primary targets of most drugs and therefore are the main targets for developing new ones.

The ever-ongoing developments in data generation and data interpretation lead towards new technologies in the field of proteomics based on liquid chromatography coupled mass spectrometry (LCMS) [88]. Nevertheless, important hurdles must be overcome at every stage of the process, from sample preparation through data analysis. One major drawback of proteomics is the lack of an amplification method equivalent to the polymerase chain reaction (PCR), for the analysis of very scarce proteins. The complexity of the proteome and the broad dynamic range of protein concentrations hamper adequate protein mapping and quantification, considering that the typical dynamic range of LCMS detection is 4-6 orders of magnitude, while in biological samples the protein concentration can span 12 orders of magnitude (from mg/ml to fg/ml). For this and other reasons, complete coverage of the proteome cannot yet be obtained and information about splice variants, isoforms and PTMs is often missing [87,88]. Technology gradually moves past these technical limitations, but the transition from bench to bedside remains challenging. Multiple MS strategies are used these days to unravel molecular mechanisms, but the applications are mainly limited to the discovery phase. Next to this so called discovery proteomics, targeted proteomics - in which researchers focus their experiments on the subset

of proteins important to their line of inquiry - has been steadily gaining attraction over the last few years [89]. In targeted proteomics experiments, selected reaction monitoring acquisition has the ability to increase the sensitivity and specificity of the analysis, also in complex samples. Finally, new initiatives are essential to promote the efficient sharing and integration of the proteomics data. In another 15 years it might thus be that cell biology students will have to be as familiar with MS as they currently are with microscopes (Figure 2.1) [90].

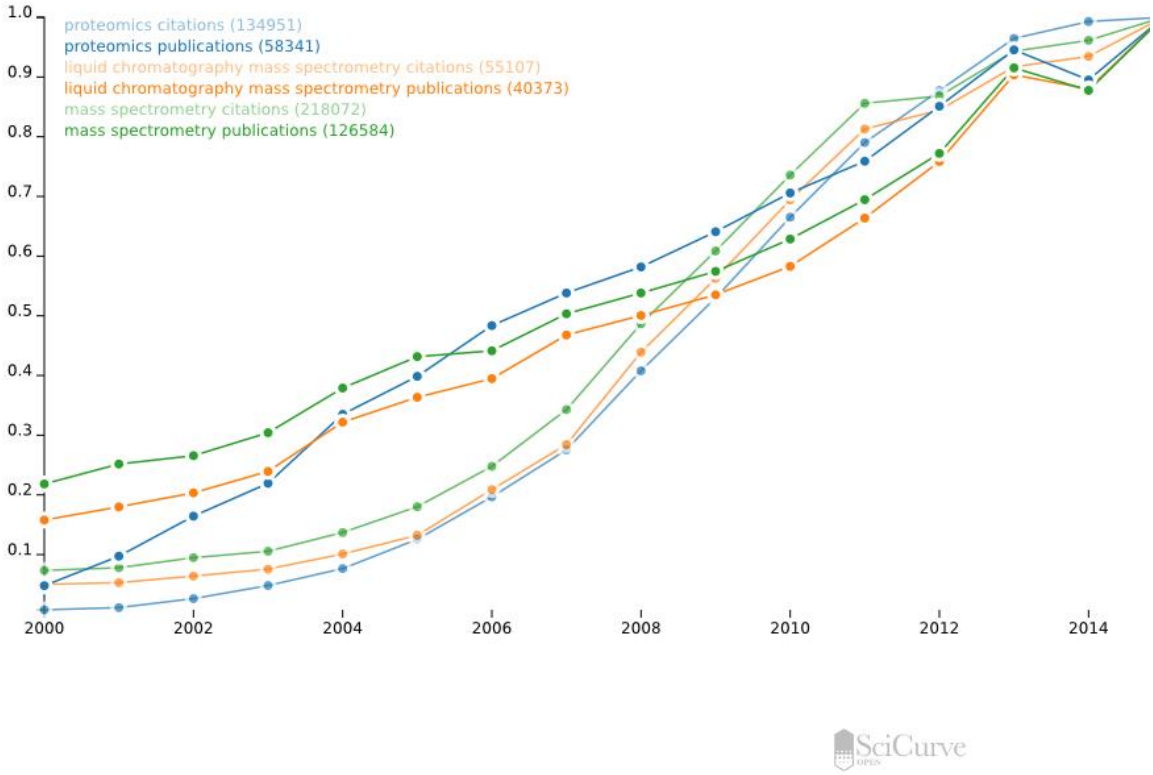


Figure 2.1 Normalized trend in publication (dark color) and citation numbers (light color) for the topics 'proteomics' (blue), 'MS' (green) and 'LCMS' (orange) over the past 15 years [91]

2.2. Strategies for protein separation

MS driven proteomics requires methods for the separation of protein mixtures into their individual components, and this for two major reasons. First, during most experiments proteins are obtained via cell lysis or out of complex biological samples, such as tissues. In an eukaryotic cell at least 50.000 – 100.000 different proteoforms may easily be present when taking PTMs into account, resulting in an enormous complexity at the proteome level. Second, most mass spectrometers can only analyze a limited amount of peptides co-eluting from the LC system. Currently, no LCMS platform can handle such complexity.

Selective methods aim to isolate individual proteins by exploiting their very specific properties such as binding specificity or biochemical function, e.g. affinity based chromatography to enrich for a certain PTM. Even though these methods are definitely very useful, the following paragraphs will mainly focus on nonselective isolation methods used in this work: sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and LC. These methods aim to fractionate a complex protein mixture in such a manner that all individual proteins and peptides derived from them, or at least a substantial fraction thereof, are readily available for subsequent analysis. The underlying principle is always the application of physical or chemical differences between proteins which will lead to different behavior in a particular environment. The two major requirements for such separation strategies are the need for high resolution and high throughput.

2.2.1. **SDS-PAGE**

The purpose of SDS-PAGE is to separate proteins according to their size, and thus molecular weight, irrespective of charge. This is accomplished by the addition of SDS to the sample, prior to the separation process in a polyacrylamide gel. Due to the exposure of denatured proteins to the negatively charged SDS, the latter will bind stoichiometrically with the polypeptide backbone. This implies that larger proteins will bind more SDS than smaller ones. As a result all protein-SDS- complexes have the same charge density and the relative differences in mass between proteins are maintained. The actual separation takes place by using an electric field to pull the proteins through the gel. The gel enhances the size-dependent separation by sieving the proteins as they migrate through it. The speed at which this takes place is dependent on the size of the protein itself as well as the gel polyacrylamide concentration. The latter determines the pore size of the gel: the higher the gel concentration, the smaller the gel pores will be, and thus the harder it will be for (larger) proteins to move through. Smaller proteins will migrate faster through the gel pores. In order to estimate the mass of the proteins in the sample, a series of protein markers with known masses can be included in one of the lanes.

2.2.2. **LC**

Chromatography is a separation technique that was invented in 1906 by a Russian-Italian botanist during his efforts to separate plant pigment, hence the name chromatography (Greek: *chroma* = color; *grafien* = writing). Nowadays, chromatography is known as any of the various techniques available for the qualitative or quantitative separation of a mixture's components, characterized by the use of a fixed stationary phase and a free moving mobile phase. Chromatography comes in all different sorts and types, including paper chromatography, thin layer chromatography, gas chromatography and liquid chromatography. They all depend on the same underlying principle: the sample mixture is dissolved in a fluid called the mobile phase. As the mobile phase moves over the stationary phase, the components

can interact with both the solvent molecules and those of the stationary matrix. The separation is then based on their differing affinity for each phase. Molecules with the least interaction with the stationary phase move the quickest, since they tend to stay in the mobile phase, and vice versa (Figure 2.2).

In proteomics, LC is an indispensable tool because of its high versatility, high throughput, high resolution and compatibility with MS. Moreover, unlike gel electrophoresis, LC is suitable for the separation of both peptides and proteins. Another advantage of LC is the broad selection of stationary and mobile phases: alternative LC methods can be based on different separation principles such as size, affinity, charge or hydrophobicity. In this work reversed phase LC (RP-LC) was used, whereby peptides are separated by their hydrophobic interaction with the non-polar groups bound on a stationary phase such as C_4 or C_{18} alkyl groups. The mobile phase composition is usually water, or a water-miscible organic sovent (methanol, acetonitrile) which is forced through the column under high pressure (RP-HPLC). Gradient elution is achieved by gradually increasing the amount of organic modifier in the mobile phase, which promotes desorption of the least hydrophobic molecules first. More hydrophobic peptides are eluted slower than the hydrophilic ones.

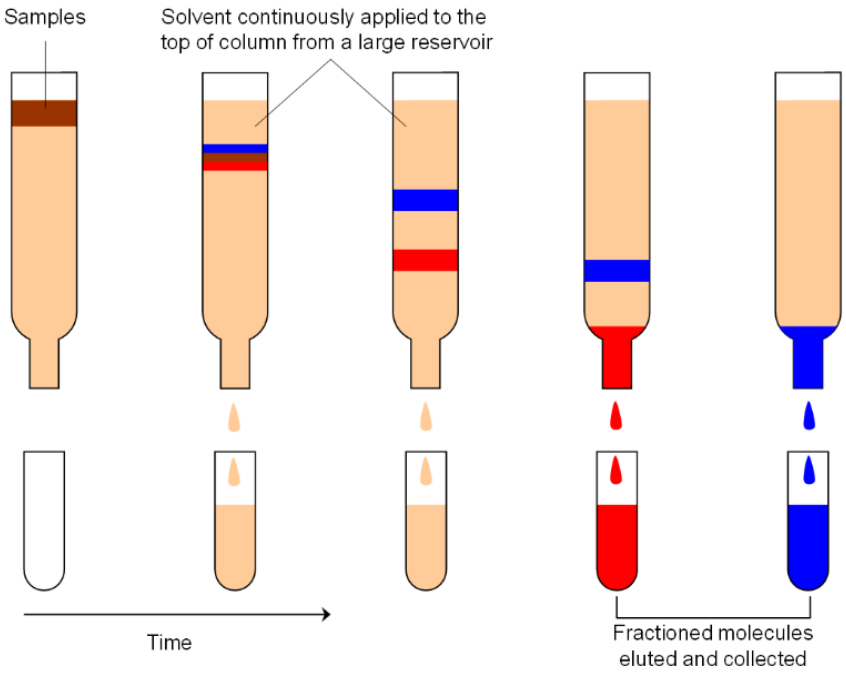


Figure 2.2 Protein purification by column chromatography
A sample (brown) is fractionated by means of column chromatography. The red fraction shows the highest affinity for the mobile phase and is eluted first. The blue component on the other hand is retained longer on column due to stronger interactions with the stationary phase.

2.3. MS based proteomics

A mass spectrometer is an analytical instrument that discriminates ions in the gas phase, based on their mass to charge ratio (m/z), with high sensitivity and specificity. This thesis will focus on the use of MS to investigate the proteome (Figure 2.3), more specific the histone proteome. MS is also applied in other fields, such as metabolomics, glycomics, lipidomics and small molecule analysis. One of the most popular hybrid designs combines two important mass analyzers: a quadrupole (Q) coupled to a time-of-flight (TOF) mass analyzer. The different units of such a Q-TOF instrument will be discussed below.

Briefly, during tandem MS the m/z ratio of a peptide is measured, whereupon it gets separated from all other ions by a mass analyzer. These targeted peptides then become fragmented in a collision cell and the so-generated fragment ions are measured by a detector. This paragraph will also briefly discuss the different methodologies for data interpretation, followed by an overview of the strategies available for quantitative proteomics.

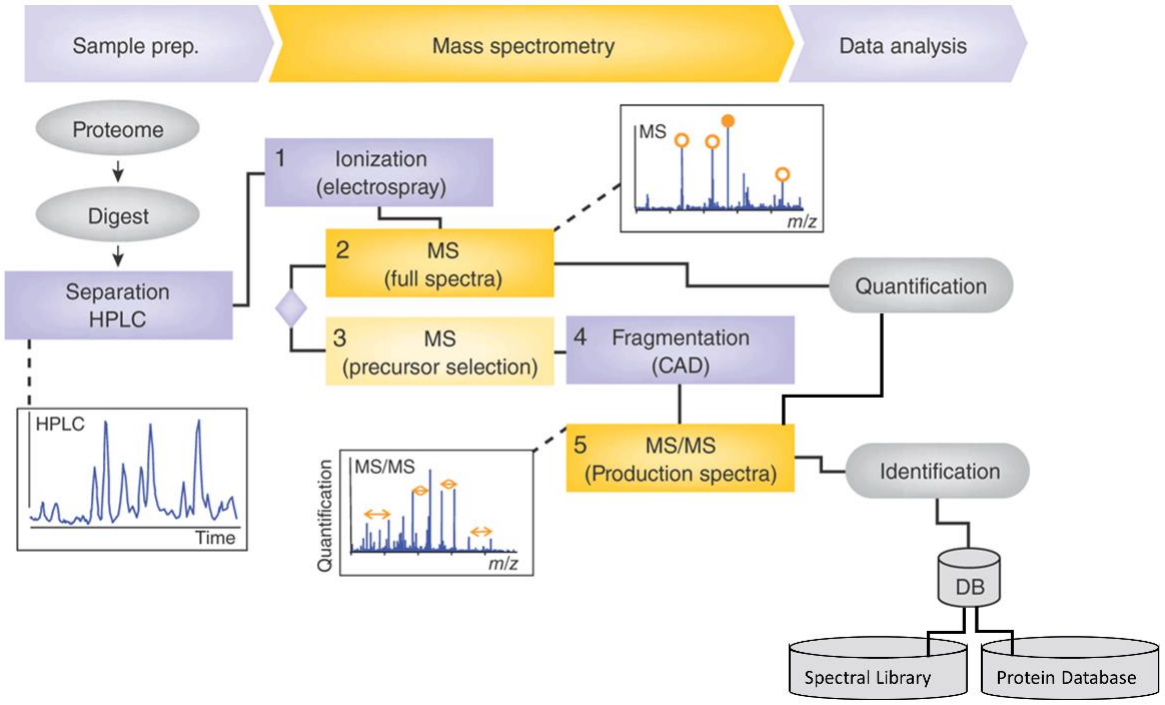


Figure 2.3 A typical proteomics experiment [92]**.**
After the digestion of proteins into peptides, the complex mixture gets separated by HPLC prior to MS analysis. Briefly, peptides are ionized using ESI, all peptides are acquired in a full spectrum MS scan, followed by the selection of a specific precursor to be fragmented and generate a MSMS spectrum. The data are then processed to be identified, quantified or both.

2.3.1. **Ionization**

The development of soft ionization techniques in the late eighties made it possible to achieve ionization of peptides and other molecules without significant fragmentation; resulting in the possibility to analyze peptides via MS. Two such methods have been widely adopted in proteomics and were rewarded a Nobel Prize in 2002: electrospray ionization (ESI) and matrix assisted laser desorption ionization (MALDI). Whereas MALDI-MS is a very sensitive method, it is mainly used for simple peptide mixtures, given that it cannot be easily coupled to an LC instrument. LC ESI-MS is more suited for the analysis of complex peptide mixtures and will therefore be discussed in more detail below.

ESI was introduced by Fenn in 1984 [93], which has transformed the field of MS ever-since. In the mid-nineties Wilm et al. introduced nano ESI and thereby allowed for the characterization of low concentration peptides due to improved sensitivity [94]. Here, peptides are dissolved in an aqueous/organic solution and forced through a narrow needle, held at high voltage (1500-3500 V). A fine spray of charged droplets emerges from the needle tip and is directed into the vacuum chamber of the mass spectrometer through a small orfice. A drying gas (often nitrogen) facilitates the evaporation of the solvent and thereby reduces the size of the droplets. The imbalance between surface tension and electrostatic repulsion finally causes a Coulomb explosion generating the ions in the gas phase (Figure 2.4). These multiply charged peptides are then drawn into the vacuum of the mass spectrometer, forming an ion beam.

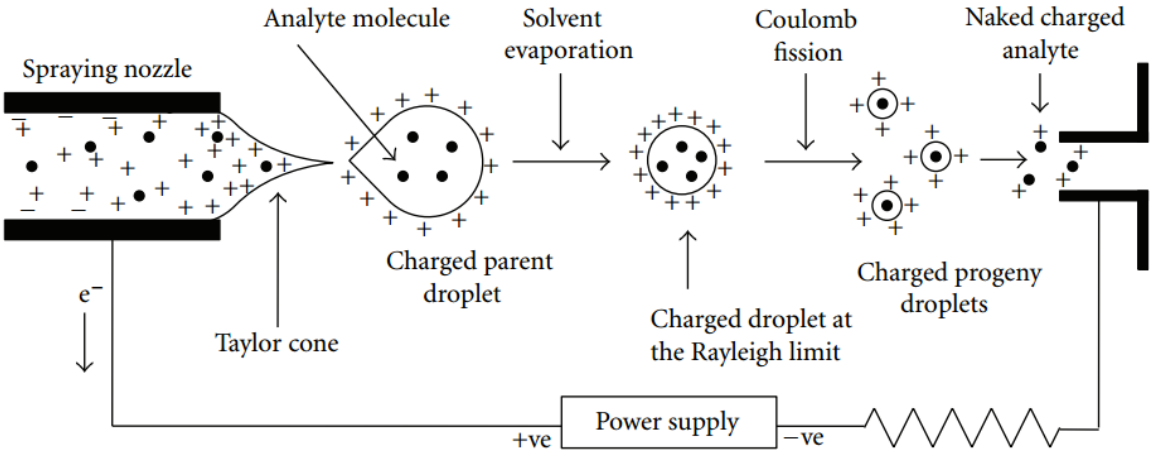


Figure 2.4 Schematic representation of the electrospray ionization process [95]

2.3.2. **Mass analyzers**

The mass analyzers are the beating heart of a mass spectrometer since they can both separate as well as analyze the ions based on their m/z ratio. In a Q-TOF instrument the Q can be used as a lens or a selector, depending on the acquisition mode, and the TOF is the actual mass analyzer (Figure 2.5).

A quadrupole consists out of four parallel metal rods: one pair with a positive electric potential and one pair with a negative potential. By alternating the radiofrequency voltage set in the two pairs of rods, ions are attracted in an alternating way to the active electrode. Since the two dimensional oscillation is dependent on the m/z ratio, only ions with a specific m/z ratio will get a stable directory and thus be able to pass through the Q. Changing the voltages alters which m/z ratio can get through without crashing into one of the rods. Quadrupoles are easy-to-use, cover a large mass range and are thus ideally suited to act as a mass filter, lens or analyzer.

Unlike Q instruments, no electric field is required to separate ions in a TOF instrument. It exploits the fact that in any mixture of ions carrying the same charge, light ions will be able to travel faster compared to the heavier ones. As the distance of the ions from pusher to detector is fixed, it measures the flight time taken by an ion, and thus its velocity, and accordingly its m/z ratio. Crucial however is that all ions possess an identical kinetic energy at the onset of the measurement, since this also affects the time-of-flight. To compensate for the variation in kinetic energy a reflectron is often used at the end of the field-free region to reflect and focus incoming ions with slightly different kinetic energies.

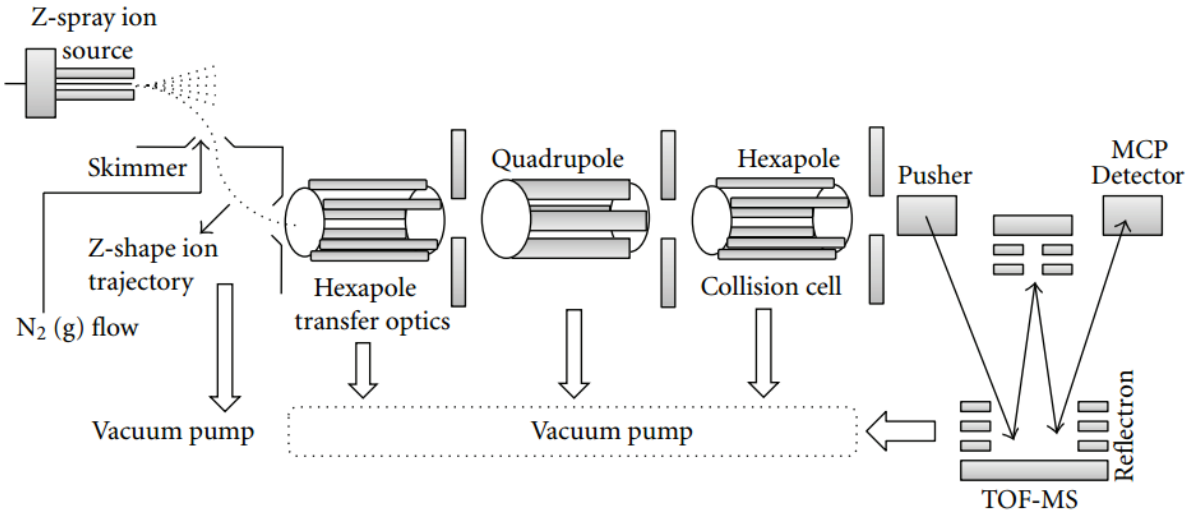


Figure 2.5 Schematic of the Waters Micromass Q-TOF Ultima ESI-MS (z-spray) [95]

Throughout this work two ESI-Q-TOF mass spectrometers were used: TripleTOF 5600 (Sciex) and Synapt G2Si (Waters) (Figure 2.6). Both mass spectrometers are state of the art technology and display some

unique features. The TripleTOF 5600 system combines high-sensitivity detection and high resolution with fast acquisition speeds, and stable mass accuracy. Moreover, the instrument enables $MSMS^{ALL}$ with SWATH acquisition [96]. In the Waters' Synapt G2Si, the quadrupole is supplemented by traveling wave technology (T-wave) which allows ion mobility separation. This separartion technique separates peaks according to their collisional cross section, specified by the charge, size and shape of the peptide. For example, peptide ions with multiple charge states can be separated as peptide ions with a higher charge state experience a higher electric field, resulting in a higher drift velocity and lower drift time. Along with the TripleTOF, this instrument can also perform data independent acquisition by means of $HDMS^{E}$ [97].

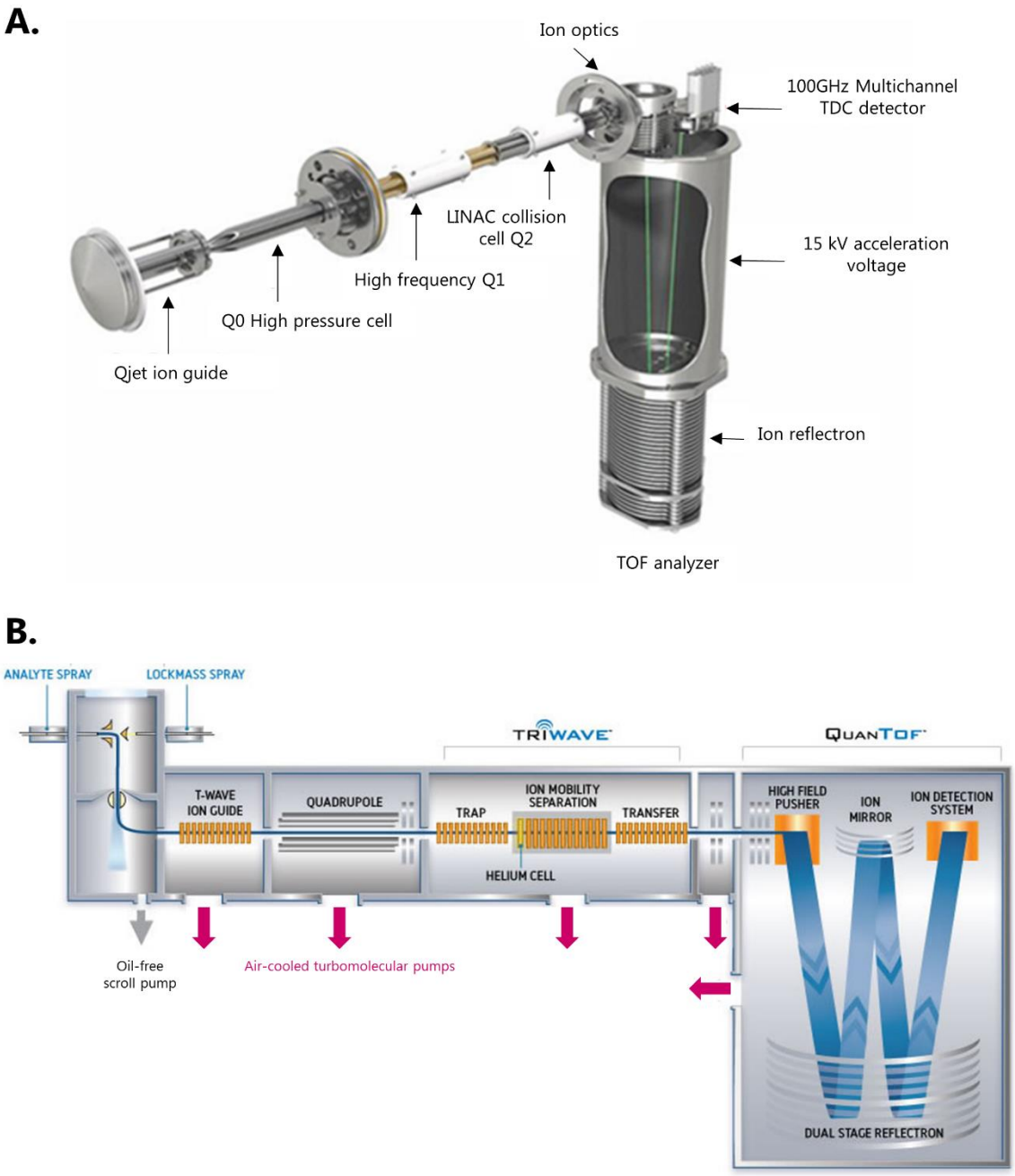


Figure 2.6 Representation of the mass spectrometers available in the Laboratory of Pharmaceutical Biotechnology: (A) TripleTOF 5600 and (B) Synapt G2Si [96,98]

2.3.3. **Fragmentation of peptide ions**

By fragmenting peptides, additional information is provided concerning their sequence and PTMs. During collision induced dissociation (CID), an intact peptide ion will be fragmented by colliding them with a stream of inert gas molecules such as argon. The generated fragments will be measured by the second mass analyzer and finally be measured by the detector.

The most informative fragments are those in which fragmentation has occurred along the peptide backbone. Two types of peptide fragments arise, each subdivided in three different classes (Figure 2.7.A): the a, b and c ions are fragments that retain the original N-terminus; while the x, y and z ions are fragments that retain the peptide C-terminus. In CID-mediated tandem MS (MSMS), interpretation generally involves the arrangement of b- or y-series ions in order of increasing mass (Figure 2.7.B). The y-ion series is often most informative because the charge most often remains on the C-terminal basic residue in a typical proteomics experiment wherein trypsin is used as the proteolytic enzyme to generate the peptides.

Electron-capture dissociation (ECD) and Electron-transfer dissociation (ETD) both induce complementary fragmentation of multiple protonated peptides by transferring electrons to them. This addition of an electron creates a radical intermediate that almost immediately decays into fragments. The observed fragment ion types for ECD and ETD are predominantly c and z ions [99]. These fragmentation strategies provide an alternative sequence coverage and are useful in PTM research and top-down proteomics.

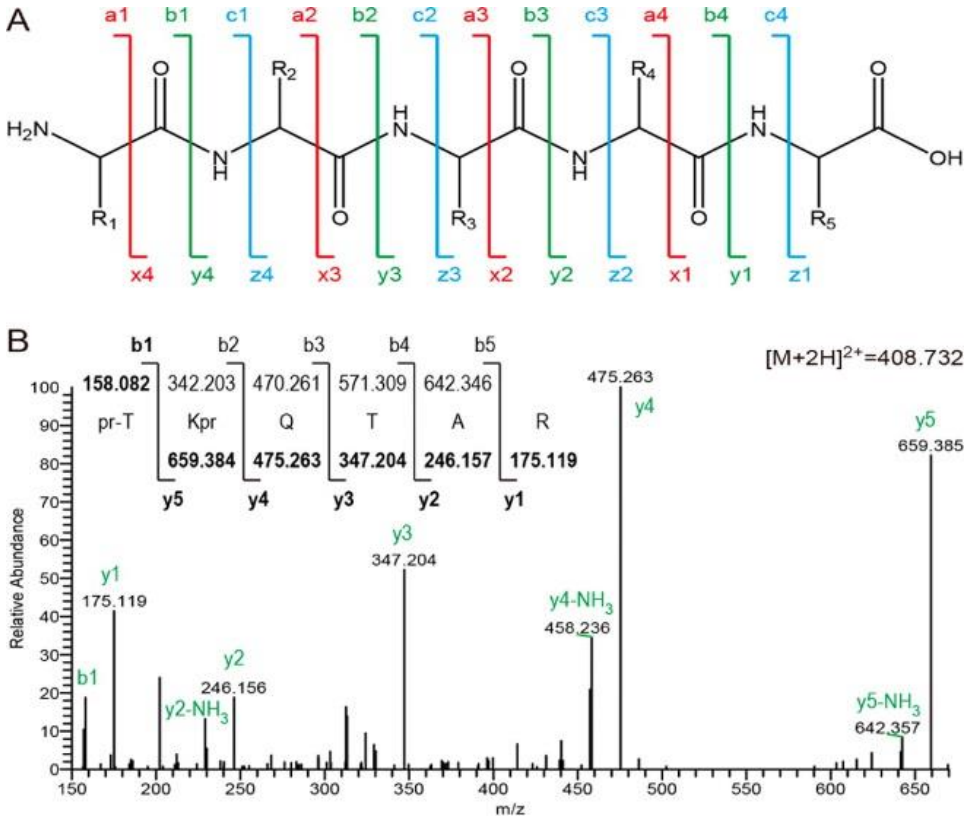


Figure 2.7 MSMS for peptide sequencing and PTM detection [66,100,101]
A: The Biemann nomenclature for peptide fragment ions. **B:** An MSMS spectrum example obtained by CID.

2.3.4. **Detection**

The final element of a mass spectrometer is the detector. Both the electron multiplier and the microchannel plate detector detect the current produced when an ion hits the surface. To increase the signal from a single electron, it gets amplified via secondary emission.

Accurate quantitative results can only be obtained when working within the linear dynamic range of any given peptide, respectively. This linear dynamic range is the range over which the ion signal is directly proportional to the analyte concentration (Figure 2.8). The linear dynamic range as well as its corresponding limits of quantifications are peptide- and mass spectrometer dependent. The typical range is situated between 2 and 5 orders of magnitude [102].

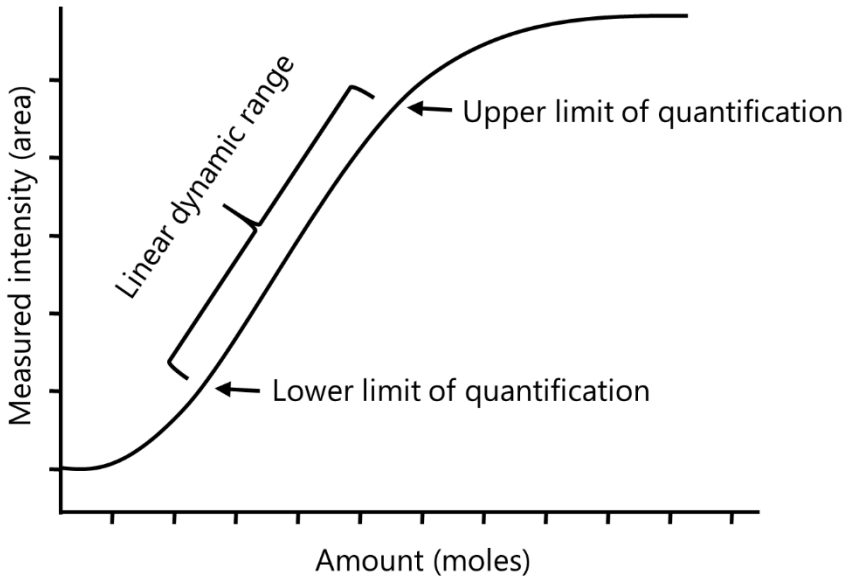


Figure 2.8 Linear dynamic range and limits of quantification

2.3.5. **Peptide identification**

All previous steps lead to the fragmentation of a precursor peptide and eventually result in a measured precursor ion m/z, a list of fragment ion m/z and intensities: the MSMS spectrum. Interpreting this spectrum and translating it into a peptide identification is the main goal in proteomics since it provides the main results of the experiment. Besides, peptide identification will also affect downstream data analysis such as quantification. Wrong peptide identifications will obviously lead to false biological conclusions. Four distinct strategies are commonly used in proteomics experiments for the identification of MSMS spectra: **(i)** de novo sequencing, **(ii)** sequence tag based approaches, **(iii)** spectral library searching and **(iv)** protein database searching.

De novo sequencing as well as sequence tag approaches aim to infer a full peptide sequence directly from the MSMS spectrum. De novo sequencing is based on the distance between subsequent fragment ion peaks, while sequence tag based approaches generate only a list of small sequences from the spectrum. If both the sequence tag and the precursor mass match in a database, the peptide is identified. During spectral database searching, spectra are compared to a spectral library of previously recorded and identified MSMS spectra. Protein database search algorithms on the other hand will use a database of theoretically created MSMS spectra for comparison. Figure 2.9 below gives an overview of the different steps that occur in these different peptide identification strategies.

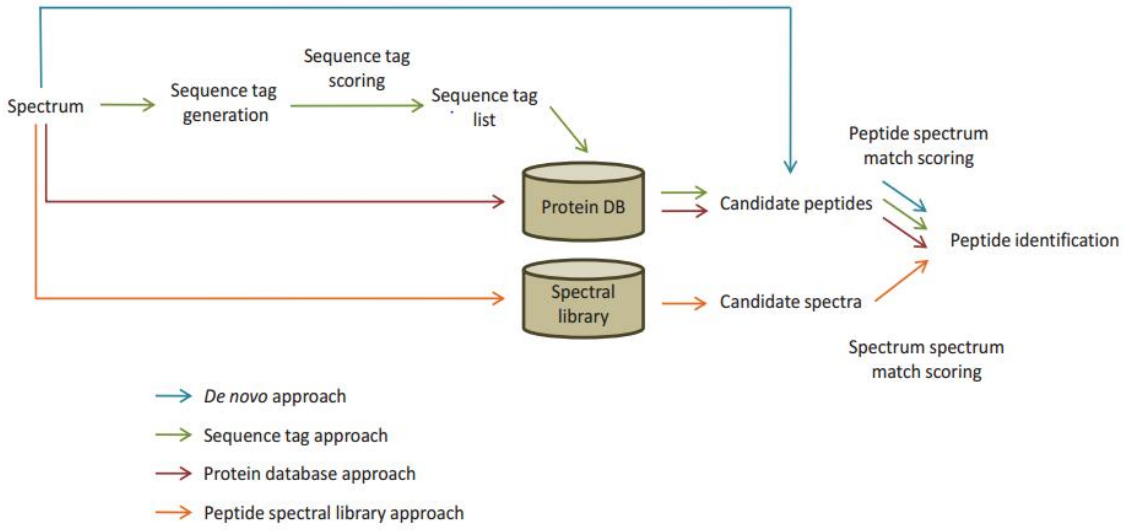


Figure 2.9 Overview of the various steps that occur in different peptide identification approaches [103]

2.3.6. **Quantitative MS proteomics**

Sometimes it is enough to know what exactly is in your sample, but more often researchers are interested in the exact quantification of a specific protein or peptide pool. This can be in absolute numbers, or relative to another sample. Since dynamic changes of histone marks are associated with many cellular processes such as transcription and DNA damage, it becomes important to quantify the changes in histone marks between two or more samples. Figure 2.10 gives an overview of all different methods and the ones used most frequently are discussed in more detail below.

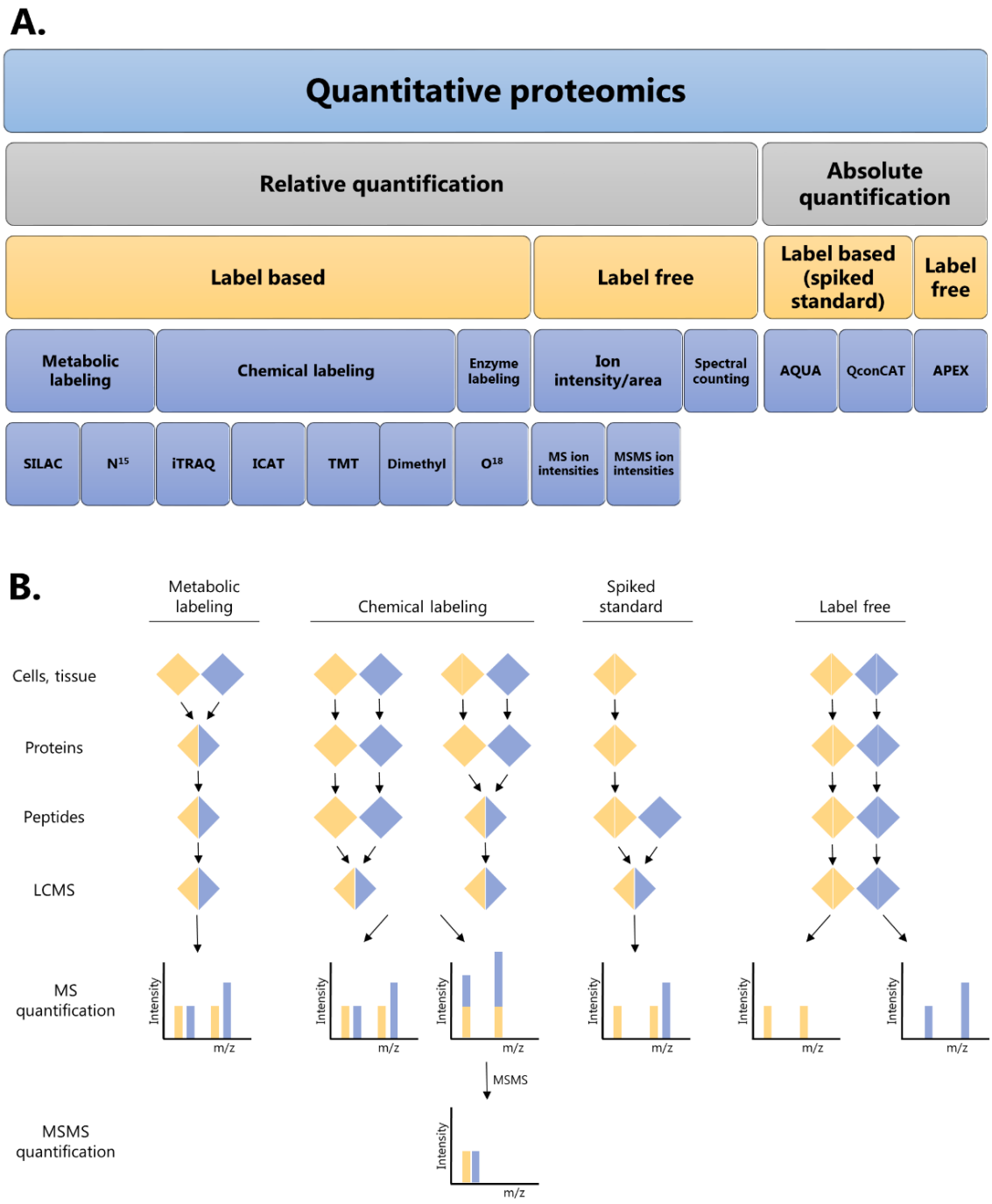


Figure 2.10 Quantitative proteomics: an overview
A: Chart representing the different strategies for quantitative proteomics. **B**: Workflows of the main quantification methods. Boxes in yellow and blue represent two different experimental conditions.

2.3.6.1. **Metabolic labeling**

Metabolic labeling requires the introduction of isotopic labels directly in the medium of a cell culture during cell growth and replication. Mann and co-workers introduced the most popular metabolic labeling strategy in the proteomics community: stable isotope labeling by amino acids in cell culture

(SILAC) [104]. This approach uses heavy labeled essential amino acids (^{13}C and ^{15}N), usually R and/or K, which eventually become fully incorporated into the proteome. A second cell population is grown in parallel in normal medium and both samples are then analyzed in the same MS run. Relative quantification is performed on the MS level by comparing the ion intensities of each of the isotopic peaks corresponding to the same peptide. The main advantage of this technique is that labeled and unlabeled samples can be combined prior to sample processing, thereby reducing technical variation. The main disadvantage is that the heavy Aa themselves can get metabolized into other Aa, impairing accurate quantification [105,106]. In 2010, Geiger et al. introduced super SILAC, which is a mixture of several cell lines that serves as an internal spike-in standard [107].

Although metabolic labeling is mostly used for cultured samples, incorporating heavy labeled Aa has also been extended to whole model organisms, such as drosophila [108], chicken [109] and mice [110]. Yet, the high cost of the diet and the extensive time required for complete labeling limits its widespread *in vivo* application.

2.3.6.2. **Chemical and enzymatic labeling**

While SILAC provides a powerful experimental strategy, metabolic labeling is not always a feasible option, for example when working with clinical samples. One way of overcoming this problem is chemical or enzymatic labeling. In general, protein extracts from the different samples that ought to be compared, are chemically labeled during or post digestion and subsequently analyzed in the same MS run. Quantification can occur at the MS (ICAT, dimethyl labeling, ^{18}O labeling) as well as the MSMS level (iTRAQ, TMT), and the main possible techniques are discussed hereafter.

Isotope-coded affinity tags (ICAT) reagents consist of three functional elements: a reactive thiol group for derivatising the reduced cysteine residues, an isotopically coded linker and an affinity tag to enrich the labeled peptides [111]. A major disadvantage however of the ICAT technology is the fact that it limits quantification to cysteine containing peptides, which is one of the least abundant Aa in the proteome. Dimethyl labeling is an alternative method, based on the chemical labeling of primary amines (peptide N-termini and the ε-amino group of K residues) with light or heavy (^{13}C, ^{2}H) formaldehyde [112], thereby labeling every peptide after trypsin digestion. Even though dimethyl labeling is probably the least expensive and most straight-forward stable isotope labeling strategy it has not been extensively used and validated in the proteomics community. This might be partly due to the fact that deuterated peptides show a small but significant shift in retention time in RP-LC compared to their non-deuterated counterparts which can cause substantial quantification errors in quantitative proteomics [113]. Alternatively, the stable heavy isotope label can be introduced into the peptide during enzymatic

digestion and the accompanying reaction with $H_2{}^{18}O$ [114]. Proteolysis results in the incorporation of up to two ^{18}O atom into the C-terminus of each proteolytically generated peptide. Although ^{18}O labeling has the potential of being the ideal chemical labeling technique, it has not become a preferred method. Incomplete incorporation and the overlap of the isotopic envelope resulting from the small mass shift that is introduced are the most likely reasons for this.

The strategies previously mentioned are all based on the mass difference generated between differentially labeled peptides. In order not to increase the complexity of the MS spectra too much, many experiment setups are limited to binary (2-plex) or ternary (3-plex) set of reagents. Yet, the use of isobaric tags for relative and absolute quantification (iTRAQ) [115] and tandem mass tags (TMT) based methods [116] allows for the relative quantification of multiple samples at the tandem MS stage. In these techniques the samples are first digested, whereupon both N-termini and lysine side chains of peptides are labeled with a different isobaric mass reagent in such a way that all derivatized peptides are isobaric and chromatographically indistinguishable. However, when subjected to CID, the different reporter ions typical for each label are generated. The ratios of the different tags at the MSMS level are then used as a quantitative read-out for the relative peptide abundance and thus serve as a surrogate measurement for the protein from which they derive. The main advantages lie in the possibility to do a multiplex analyses in the same LCMS run, thereby limiting the effect of technical variability. However, challenges still remain, such as the underestimation of the fold change and the inherent problems with precision and accuracy [117–119]. As for today isobaric labels are still a valued technique, but the high cost in combination with these drawbacks pave the way for label free quantitation strategies.

2.3.6.3. **Label free quantitation**

Label free quantification does not require expensive labelling, nor the additional sample preparation steps that can result in experimental variability. Furthermore, the technique has become more reliable as a result of the constantly improving LC-MS equipment. This quantification strategy is based on the observation that the electrospray process gives rise to a signal response that is linearly correlated with the increasing concentration of the peptide. Label free approaches can be divided into two main groups by the way the abundance of a peptide is measured: **(i)** spectral counting and **(ii)** peak measurements. In the first group the total number of peptide spectra matching is taken as an indicator of the protein abundance [120]. Yet, this may lead to a bias in favor of large and more abundant proteins. Most label free analysis methods however rely on the direct evaluation of peak measurements (ion intensity and peak area) for relative quantification. The LCMS analysis of samples to be compared is then carried out separately and only at the final stage LC chromatograms are overlaid and compared at the MS or MSMS level. Although this method is relatively straightforward, several considerations must be taken into

account to ensure accurate and reproducible quantitation. Potential sources of technical variation (and thus experimental error) are unequal sample loading and run-to-run variation in the LCMS process, amongst others.

Label free quantitation by measurement of peptide fragment signal intensity at the MSMS level was rarely used up to now due to the stochastic nature of data dependent acquisition (DDA). However, data independent acquisition (DIA) has the potential to make large scale MSMS quantitation a more viable alternative to label free MS quantitation. In 2012 an approach was developed, named "Sequential Window Acquisition of all THeoretical Mass Spectra" (SWATH) [121], whereby complex mass spectra of a predefined mass range are queried for the presence of specific peptides. The latter is done by using libraries of qualified peptide spectra, in combination with their accompanying retention time. In the second approach, named MS^E, the collision energy is alternated between low and elevated energy scans. In the low energy scans the intact precursors are measured, in the corresponding high energy scan all these peptide ions are subjected to CID without isolation of a specific precursor. Afterwards it is possible to align the precursor and the corresponding fragments for identification as well as quantification [122,123].

2.3.6.4. **Spiked standard**

In 2003, Gygi and his team introduced an innovative strategy used for targeted quantitative proteomic analyses: Protein Absolute Quantitation (AQUA) [124]. In contrast to the previously described methods, AQUA enables absolute protein quantitation instead of relative quantification, using stable isotope labeled peptides (mostly ^{13}C and ^{15}N). An AQUA peptide, which is a synthetic peptide corresponding to a peptide of interest, is spiked into a sample containing the native peptide. Both peptides will behave exactly the same during the following sample preparation and LCMS process (elution, ionization) and thus remain indistinguishable until the MS level where quantification becomes possible by comparing the signal intensity of both the spikes and sample peptide.

2.4. MS for histone analysis

Over the last decades, histone modification analysis has been mainly conducted using site specific antibody-based techniques such as Western blotting, immunofluorescence and chromatin immunoprecipitation, among others [125]. While these techniques are very useful, drawbacks such as highs costs, cross-reactivity, epitope occlusion and the need for prior knowledge make the discovery of new PTMs difficult. Hence, it is no surprise that the significant increases in mass accuracy and resolution of mass spectrometers have brought these instruments to the forefront as an analytical strategy to detect histone PTMs (hPTMs).

The same procedure as described in section 2.3 can be used for identification and localization of PTM sites. Each PTM will alter the molecular weight of the Aa it is located on. This can then be taken into account while interpreting the MS and MSMS spectra for identification. For example acetylation on lysine will result in a mass increase of 42.0106 Da, compared to an unmodified lysine reside. Nevertheless, the physical properties of histones, together with the large number of PTMs and variants make identification of histone PTMs a laborious task. The major difficulties arise prior to and post MS acquisition: during sample preparation and data analysis. Both these topics will be discussed below. First, we focus on three major strategies for sample preparation (Figure 2.11) namely bottom-up, middle down and top down approaches, and draw up some pros and cons for each strategy (section 2.4.1 till 2.4.3). Next, we provide some insights into the hurdles during interpretation of the MSMS spectra (section 2.4.4). Finally, since MS methods enable simultaneous detection and quantification, we elaborate further on the latter in section 2.4.5.

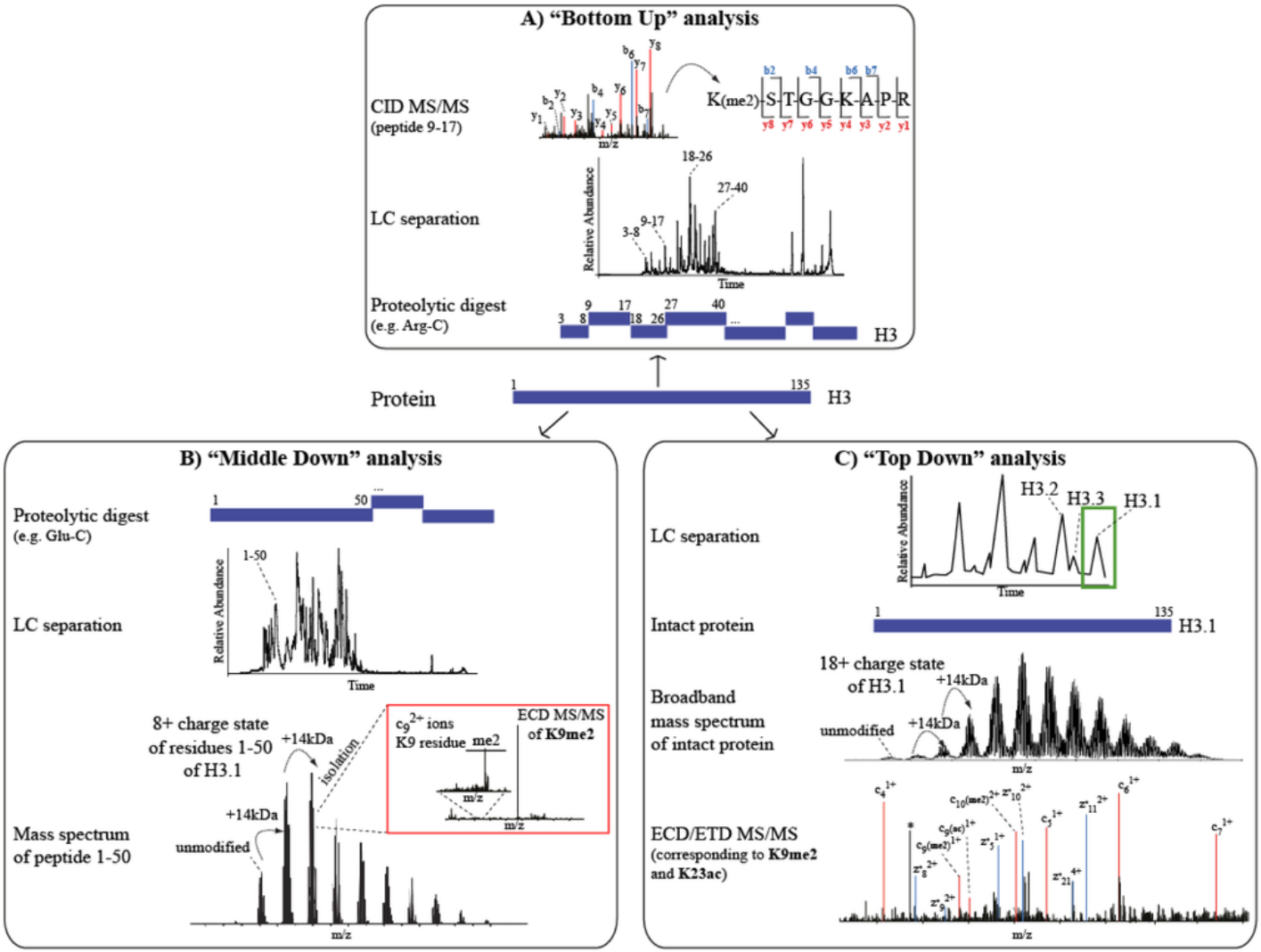


Figure 2.11 Comparison of peptide centric versus protein centric MS strategies for the analysis of histone PTMs [126]

A: During bottom-up analysis, histones are digested by a protease such as Arg-C or chemically derivatized prior to trypsin digest. The resulting peptides are then subjected to LCMS analysis using CID. This fragmentation technique generates b- and y-ions, which permits for the exact localization of a PTM. **B:** The middle down approach is also a peptide centric approach, but in contrast with the bottom-up approach other enzymes (such as Glu-C) are used for digestion. This results in larger peptides, compared to bottom-up. In the illustrated example, the full MS spectrum corresponding to peptide 1–50 is reported. The peak corresponding to 8+ charge state is then isolated and subjected to Electron Capture Dissociation (ECD) fragmentation. **C:** In the top down approach, a purified intact histone is analyzed in MS mode. In the example, the modified form of H3.1 is reported (middle panel) and the zoomed region of the ECD spectrum, corresponding to 18+ charge state of H3.1, is shown (bottom panel).

2.4.1. **Bottom-up histone analysis**

Bottom-up is the most widespread strategy for both identification and quantification of histone PTMs (Figure 2.11). Since the N-terminus of histone proteins are highly enriched in basic residues like lysine and arginine a regular bottom-up MS approach is a challenge. Standard proteases such as trypsin will cleave after every basic residue and therefore generate small hydrophilic peptides with insufficient chromatographic retention. Moreover, the presence of frequent modifications on lysine residus will reduce digestion efficiency. Two approaches can be used to overcome these problems. **(i)** The use of Arg-C enzyme or Glu-C could partly circumvent this problem, but they both have their limitations. Arg-C only cleaves at arginine but lacks efficiency and specificity resulting in poor MS analysis [127]. Glu-C on the other hand cuts at the C-termini of glutamic acid and aspartic acid residues which are less common Aa in histones, generating large and multiply charged peptides whose MSMS spectra are difficult to interpret. **(ii)** A solution can be found in chemical derivatization of histones on lysine and monomethylated lysine, followed by trypsin digestion [128]. The benefits of trypsin concerning efficiency and specificity remain, but larger peptides (6-20 Aa) ending on only arginine are formed. Moreover, the addition of a hydrophobic group on lysine increases retention time.

Several lysine derivatization strategies have been developed over the years. One of the most adopted protocols applies propionic anhydride for derivatization (Figure 2.12) [128]. Nevertheless, also other derivatization methods are used. For example, Liao et al. published a paper using the propionic acid N-hydroxysuccinimide ester instead of the anhydride [129]. The Smith group designed a method that uses deuterated acetic anhydride to acetylate all free and monomethylated lysine residues [130,131].

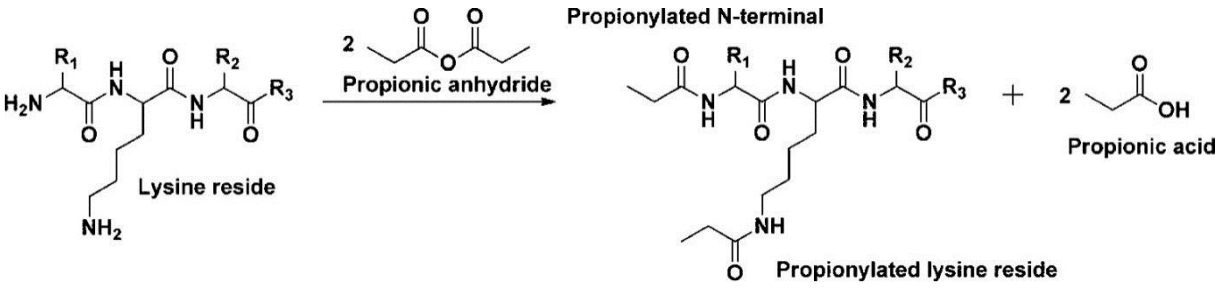


Figure 2.12 Chemical derivatization of lysine residus by propionic anhydride [66]
When propionylating histones prior to trypsin digestion, all free amine groups on the N-termini and the ε-amino group on free or endogenously monomethylated lysines will be blocked by the chemical propionyl group. When the N-terminus or a lysine residue is covered by an endogenous modification other than monomethylation, a propionyl group can no longer bind because the residue is blocked by that modification itself. Consequently, trypsin can no longer cleave after these derivatized lysine residues but results in proteolysis only C-terminal to arginine residues, mimicking an Arg-C digestion, but with the efficiency and reproducibility of trypsin. By propionylation post digestion, all newly generated N-termini will be propionylated as well.

The advantage of a bottom-up MS strategy lies in its high sensitivity and less complicated data analysis, compared to the middle down and top down approaches. Despite the many applications of bottom-up histone analysis, some limitations emerge as well. First, since bottom-up commonly uses CID for fragmentation, labile PTMs such as phosphorylation tend to be lost. Also, random backbone cleavage of the peptide backbone can be hindered due to the great amount of basic residues that prevent random protonation. Finally, because digestion results in short peptides (6-20 Aa) connectivity between long distance PTMs is lost and it is hard to assign PTMs to specific variants ([126,132]. Alternatively, to overcome these drawbacks, larger histone domains or intact histone proteins can be analyzed with respectively the middle down or top down strategy [133]. Pros and cons are listed in Table 2.1.

2.4.2. **Middle down histone analysis**

As the name implies, the middle down approach is a compromise between bottom-up and top down strategies (Figure 2.11). It is still a peptide centric strategy, but it differs in the size of the peptides generated (>2 kDa). To obtain these larger peptides a protease other than trypsin must be used. In fact, AspN and GluC are often used since they cleave less frequently occurring Aa [126,134,135].

The main advantage of this strategy is the possibility to examine coexistence of multiple PTMs. For example, when using Glu-C the entire H3 N-tail is cleaved since the first glutamic acid is at position 50. This N-terminal part contains most PTMs decorating this histone, which can then be related to each other. Moreover, ETD as a fragmentation strategy is better suited to preserve labile PTMs than CID and results in even backbone fragmentation. On the other hand, the fragmentation efficiency of this fragmentation technique is low. Another caveat of middle down MS is the decreased sensitivity (relative to bottom-up analysis), even though it is still remarkably higher than top down approaches. This can be explained by the fact that the MS signal is diluted by the different charge states of a larger peptide ion on the one hand, and by the increasing number of possible PTM combinations on the other. Furthermore, there is a need for specialized software to identify the complex combinatorial networks of histone PTMs and the algorithms used herefor are still in their infancy (Table 2.1) [136–138].

2.4.3. **Top down histone analysis**

Histones are rather small proteins (11-21 kDa) and therefore very suitable for top down MS analysis [139]. There is no digestion step required in this technique and complete proteins can be analyzed by the mass spectrometer (Figure 2.11). The primary advantage lies in the fact that all PTM crosslinks remain available for analysis, and thus can provide information about the stoichiometry of all modifications. Also, variant specific PTMs can be distinguished more easily. Yet, the same drawbacks as for middle down analysis remain: incomplete fragmentation, sensitivity drops, prefractionation is often required to

enable analysis of all the different proteoforms and instrument as well as software requirements are more advanced in terms of mass resolution and mass accuracy to deconvolute the very complex MSMS spectra (Table 2.1) [66,126,132,134,135].

Table 2.1 Pros and cons of different MS strategies for histone analysis [134]

MS strategy	Scope	Advantages	Disadvantages
Bottom-up	Small peptide fragments	Best sensitivity	Lose connectivity of most PTMS
		Easiest analysis	Generally paired with CID
			Labile PTMs lost
			Non random backbone cleavage
Middle down	Medium peptide fragments (~50 Aa)	Better connectivity than bottom-up peptides	Complicated data analysis
		Better sensitivity than top down	Lose connectivity of some PTMs
		Retain labile PTMs	Less sensitive than bottom-up
		Even backbone cleavage	Low fragmentation efficiency of ETD
Top down	Entire proteins	Complete connectivity of PTMs	Difficult data analysis
		Paired with ETD	Worst sensitivity
		Retain labile PTMs	Low fragmentation efficiency of ETD
		Even backbone cleavage	

2.4.4. **Interpreting MS data during histone analysis**

Many search engines have been developed to quickly and accurately analyze large volumes of proteomics data. Popular softwares like SEQUEST, Mascot and Andromeda align experimental MSMS spectra with theoretical peptide fragmentation patterns [140,141]. Yet, identification of PTM bearing peptides, and histones in particular, is more challenging than that of the unmodified forms. This is because search alghoritms have to take into account the diversity of multiply modified forms that might exist. A database search with many PTMs will result in low search speed and an increased number of false positives. Besides, identifying new PTMs poses a challenge when performing a database search, since the new PTM will most probably induce an unknown mass shift. Finally, next to the identification of a PTM, correct localization is a struggle as well. In the paragraphs below, we discuss some of the efforts that have been made over the years in order to overcome these data analysis issues.

(i) Several alghortims have been described with a procedure called "non restricted sequence alignment" [142–146]. These algorithms enable identification of a mass shift - caused by a novel PTM or mutation - by assuming that the mass difference in the MS spectrum can be localized to certain residues through comprehensive MSMS spectra alignment. Still, these methods suffer from ambiguous alignment and noisy background which makes manual validation inevitable [66]. **(ii)** Fu and co-workers proposed a method to estimate the individual PTM false discovery rate (FDR) from the global FDR [147]. **(iii)** Yuan et al. developed a strategy, named iterative search for identifying PTMs (ISPTM), for tackling this

challenge. In this approach unmodified spectra are identified in basic search, afterwards the leftover unidentified spectra are analyzed by iterative searches of many PTMs using a small number of them (usually two) in each search [148]. **(iv)** The group of Garcia thought of a strategy to evaluate search engines on identifying histone PTMs [149]. Hereby, histone data are searched with separate modifications and filtered for confident results (FDR <1%). Next, redundant identifications between different searches with low scores are discarded in order that only confident identifications remain. The authors put some of the more well known search engines to this test and found that two search engines, pFind and Mascot, identify most of the confident results at a reasonable speed. **(v)** Last but not least, also the technical improvements in mass spectrometers can facilitate database searching. Thanks to the higher mass accuracy of MS and MSMS spectra, the number of peptide and protein candidates gets limited, which reduces the search space [150].

Even though the previous paragraph illustrates identification of PTMs is a laborious task, this is even more true for determining the exact localization thereof on the peptide backbone. To accurately map a PTM's localization, different tools have been developed. **(i)** Ascore measures the probability of correct PTM localization based on the presence and intensity of site-determining ions in MSMS spectra. While this subjective measure will vary from dataset to dataset, an Ascore of >20 is generally a "good" Ascore, with a site being localized with >99% certainty [151]. Maxquant/Anderomeda adopts a similar strategy for PTM scoring [152]. **(ii)** Another method for scoring the PTM-localization probability is the Mascot Delta Score or MD-score [153]. Here, the Mascot ion-score difference between peptide identification with different site locations are compared. The bigger the difference, the better. A MD-score of 10 means probabilities of 91% and 9% for two PTM localizations (based on the same spectrum). As the name suggests, this scoring method is only available when using Mascot as a software tool. **(iii)** In addition to these tools, many other scoring modules or softwares have been developed over the years, including: Phospho Score [154], phosphorylation localization score in Inspect [155], SLoMo [156], PhosphoRS [157], Phosphinator [158], SLIP score in protein Prospector [159], ModLS and D-score [160].

2.4.5. **Quantification of posttranslational histone modifications**

Histones are involved in many cellular processes such as transcription and DNA damage repair. The dynamic changes of PTMs play an important role herein. Hence, many research groups want to quantify changes of histone marks between two or more samples. The various strategies mentioned in section 2.3.6 can also be used for measurement of histone modifications, variants and turnover [161–169]. Since chemical labeling with dimethyl also labels primary amines, this is often replaced by isotopic propionic anhydride (e.g., H_{10} vs D_{10} or $^{12}C_6$ vs $^{13}C_6$) to label two pools of proteolytic peptides. In the following

paragraphs we will further discuss the most popular methods for determing the stoichiometry of histone marks.

(i) First is absolute quantification of modification states. For this, synthetic, isotopically labeled peptides can be used as internal standard for both relative and absolute quantification [124,170]. The peptides are synthetized with the same sequence (and PTMs) as the peptide of interest, but isotopically encoded. Briefly, relative quantification is obtained when the intensity of each native modified peptide is compared with that of the internal standard. Absolute quantification can be achieved by making a calibration curve of the ion intensity versus the concentration of the peptide standard. **(ii)** Next is isotopic labeling *in vitro* by chemical derivatization. Analysis of iTRAQ data at the peptide level surfaced e.g. an interesting aberrant proteolytic product of a histone protein: clipping of the histone H2A C-tail [171]. Nevertheless, both TMT and iTRAQ have been mainly employed on chromatin for protein level profiling, with no focus on PTM changes [172–174]. **(iii)** Further, *in vivo* metabolic labeling has emerged as a powerful tool for accurate peptide/protein quantification. Even though SILAC is preferentially used to profile protein levels, it has also been successfully applied to identify and quantify hPTMs [163,164,175,176]. Zee et al. even used the SILAC approach to measure the turnover of both hPTMs and variants [177]. A variation of SILAC, known as heavy methyl SILAC (hmSILAC), is used for identification of methylation at K and R. In this approach, "heavy" methionine is added to methionine depleted media. From there, all proteins (histone as well as non-histone) will be enzymatically methylated by a heavy methyl group. The presence of a specific light and heavy peak pair in the MS spectrum can then be used for high confidence identification of a methyl group and subsequent quantification [178]. A drawback of SILAC is that the approach is limited to a comparison of no more than three functional states in a single experiment (unlabeled, $^{13}C_6$- and $^{13}C_6$ $^{15}N_4$-labeled Aa). Also, some samples - such as clinical samples - are not amenable to metabolic labeling as they contain no actively dividing cells to incorporate the isotopic label. This limitation was overcome by Mann and co-workers through the introduction of super SILAC [107]. In this approach, a standard heavy labeled proteome mixture is spiked into clinical samples, generating a universal reference for quantification. In order to use this strategy for quantifying hPTMs, a comprehensive set of heavy labeled histone peptides (containing virtually all known hPTMs) should however be generated as a universal reference. **(iv)** Lastly, label free quantification can be used for the relative quantification of histone PTMs. This technique enables determination of dynamic changes among dozens or even hundreds of samples and is cost-effective. Currently, most label-free quantification experiments are based on the calculation of the peak areas of the modified-peptide parent ions after standard DDA acquisition. In this type of analysis, the MS signal of all different modified peptides and their corresponding unmodified counterparts are considered. Specifically, the peak areas of the unmodified and all differentially modified forms of a histone residue of interest (e.g., H3K9 residue)

are summed and designated as 100%, and an individual form of a histone peptide is calculated as a fraction of the summed peak areas [179]. Recently however, label free quantitation methods based on DIA have been developed: HDMS^E and SWATH [121,180]. The latter has been implemented by Sidoli et al. to monitor the dynamic changes in histone marks during differentiation of hESCs and mouse trophoblast stem cells. A general concern for label free quantitation is the variation in MS detection efficiency, induced by PTMs. This is partially due to the difference in ionization efficiency of peptides with the same sequence but a different PTM. This can be overcome by applying correction factors generated by spiking in synthetic histone peptide standards (internal correction) or by using information derived from independent MS analysis of those standard synthetic histone peptide mixtures (external correction). Nevertheless, due to the combinatorial versatility of hPTMs on the one hand and the high cost on the other, synthesis of all these peptides is not feasible [181].

"In the beginning there was nothing, which exploded."

— Terry Pratchett

CHAPTER 3:

OUTLINE AND AIMS OF THIS THESIS

3. OUTLINE AND AIMS OF THIS THESIS

For over a decade, scientists have had access to a reference human genome: the complete set of genes in our cells. It is hard to think of any branch of human biology that has not benefited from the elucidation of the human genome sequence, given that it forms the foundation for understanding how variation in the genetic code can affect human health. But despite the technological improvements and biological discoveries, much remains to be understood about how genetic information is used by the individual cells in our body. This is where epigenetics comes in. Operating "on top" of the genome, epigenetic mechanisms collectively help in regulating gene expression and variations thereof without altering the primary DNA sequence. Metaphorically, the many different cell types throughout the body (found within brain, bone, heart, skin, etc.) represent different readers of that same three billion letter DNA encyclopedia, each highlighting their favorite parts, dog-earing certain pages, annotating interesting paragraphs, and crossing out things they find dull or uninteresting. This thesis focuses on one specific part of epigenomics that has remained relatively understudied compared to e.g. DNA methylation: histone PTMs. Nevertheless, these modifications also participate in gene expression regulation by influencing genome accessibility and/or recruitment of other effector proteins involved in transcription. We approach this matter from a biological (Chapter 4) as well as a technical point of view (Chapter 5 and Chapter 6). The biological background was already described in **Chapter 1**; whilst **Chapter 2** summed up the most important techniques used throughout this work.

Our first aim was to gain more insight in the importance of histone PTMs and histone clipping in particular. One very relevant model to study epigenetics are human embryonic stem cells (hESC). Differentiation of these hESCs provides a unique opportunity to study the regulatory mechanisms that facilitate cellular transitions in a human context. In order to achieve this goal, several questions were put forward**: (i) "What is the biological relevance of histone clipping in hESC?" (ii) "Where does this clipping take place?" and (iii) "Which enzyme mediated this process?"**. As described in **Chapter 4**, we were able to show for the first time that histone H3 clipping takes place in hESC, in contrast to completely differentiated Raji cells that do not show any clipping. We could pinpoint several cleavage sites and characterize the enzyme responsible to be a serine protease. Nevertheless, we have no conclusive evidence for the epigenetic relevance (or biological relevance whatsoever) of this process. Many theoretical approaches can explain how H3 clipping could be involved in several processes linked to gene expression. Yet, the question still remains as to whether H3 proteolysis correlates directly to or causes such processes. In **the first part of Chapter 7** we elaborate further on this issue and put forward a proposal for categorization of histone proteolysis into complete degradation and specific clipping.

Our second objective was to improve both identification and quantification of histone PTMs through bottom-up MS. Chemical derivatization (mostly propionylation) prior to trypsin digestion is required to generate sufficient peptides of adequate length. Unfortunately, this additional step during sample preparation goes with a lot of technical variation, obscuring biological changes of epigenetic relevance. Therfore, the main research question is: **"Do current propionylation protocols meet the technical standards needed for proper identification and quantification of histone PTMs?"**. In **Chapter 5** we disclose several pitfalls in propionylation that can hinder identifcation as well as quantification and make some suggestions for streamlined data analysis. We try to tackle each of these pitfalls more in depth in **Chapter 6** and put forward a propionylation method of choice for future MS analysis. General conclusion towards different propionylation strategies can be found in **the second part of Chapter 7**.

"If you try and take a cat apart to see how it works, the first thing you have on your hands is a non-working cat."

— Douglas Adams

CHAPTER 4:

HISTONE CLIPPING IN HUMAN EMBRYONIC STEM CELLS

Adapted from "**Identification of histone H3 clipping activity in human embryonic stem cells**"

Paulien Meert*, Liesbeth Vossaert*, Ellen Scheerlinck, Pieter Glibert, Nadine Van Roy, Björn Heindryckx, Petra De Sutter, Maarten Dhaenens and Dieter Deforce

Stem Cell Research (2014) 13 (1), 123-134

* Equal contribution

4. HISTONE CLIPPING IN HUMAN EMBRYONIC STEM CELLS

4.1. Abstract

Histone PTMs are essential features in epigenetic regulatory networks. One of these modifications has remained largely understudied: regulated histone proteolysis. In analogy to the histone H3 clipping during early mouse ESC differentiation, we report for the first time that also in hESC this phenomenon takes place in the two different analyzed cell lines. Employing complementary techniques, different cleavage sites could be identified, namely A21, R26 and residue 31. The enzyme responsible for this cleavage is found to be a serine protease. The formation of cleaved H3 follows a considerably variable pattern, depending on the timeframe, culture conditions and culture media applied. Contrary to earlier findings on H3 clipping, our results disconnect the link between declining Oct4 expression and H3 cleavage.

4.2. Background

In all eukaryotes, DNA is tightly associated with histone proteins in order to form chromatin, of which the fundamental subunit is the nucleosome [36]. Each nucleosome consists of four different core histone types (H2A, H2B, H3 and H4), which have been very well evolutionarily conserved. Chromatin structure is essential for compaction of genomic DNA but also represents a physical barrier to control DNA accessibility and gene expression.

In ESC the delicate balance of self-renewal and differentiation into specific lineages is determined by many lineage-restricted promoters that are associated with highly combinatorial histone PTM patterns which may determine their selective priming of gene expression during lineage commitment. Together with DNA methylation, ATP-dependent chromatin remodeling, RNA interference, ncRNA and incorporation of histone variants, these properties form the "epigenetic signature" [55].

Not only ESC differentiation but also other biological contexts are characterized by a continuous interplay of installation and removal of histone PTMs. To accomplish the latter, several mechanisms can be at play. Apart from enzymatic elimination of modifications [55,68] and histone exchange [55,182,183], also regulated proteolytic histone cleavage has been suggested to play such role [184,185].

Duncan et al. showed that H3 is proteolytically cleaved at its N-terminus during early differentiation of mouse ESC (mESC) and they provide evidence for the regulatory capacity of covalent modifications herein [184]. Cathepsin L was found to cleave histone H3, with alanine 21 being the primary site of cleavage [184,186]. This truncated H3 form is detected during the first days of both monolayer differentiation (with and without retinoic acid (RA) induction) and embryonic body (EB) formation. Similar clipping events of H3 associated with other cellular processes including viral infection [187,188], aging [189,190] and sporulation [191] have also been reported. Additionally, protease activity towards H3 was also found in chicken liver and Tetrahymena micronuclei [192–195]. Although the molecular consequences of any histone clipping event are yet to be defined, these data seem to suggest an evolutionary conserved process.

Here, we show for the first time that histone H3 clipping also occurs in human ESC lines (hESC) in addition to mESC. Several cleavage sites were assigned and the clipping enzyme was characterized as a serine protease. The manifestation of this proteolytic event can theoretically have an impact on several levels such as pluripotency and differentiation. Our results indicate that H3 cleavage can indeed be accompanied by a loss of Oct4 expression but does not exclusively occur in this condition. Hence, a direct correlation between these two processes cannot be claimed.

4.3. Methods

Cell culture

Two hESC lines were used: the UGENT2 cell line (XX), created in-house and the WA01 Oct4-eGFP knock-in reporter cell line (XY), obtained from WiCell [196]. Both cell lines were cultured (5 % O2 and 5 % CO2 at 37 °C) on a feeder layer of Mitomycin C inactivated MEF. Cells were passaged every 4 to 6 days, using 1 % collagenase type IV and glass beads. Culture medium consisted of knock-out DMEM supplemented with 2 mM L-glutamine, 1 % non-essential Aa, 20 % knock-out serum replacement, 4 ng/ml bFGF, 100 U/ml penicillin and 100 µg/ml streptomycin.

Feeder-free cultures were maintained on a Vitronectin XF coating (Primorigen), in combination with Essential 8 medium (E8, Life Technologies). Cultures were split every 2 to 3 days by means of EDTA-passaging.

Differentiation was induced by omitting bFGF from the medium and by adding RA to the culture medium at a final concentration of 2 µM.

THP-I cells were maintained in suspension culture in completed RPMI consisting of RPMI medium supplemented with 2 mM L-glutamine, 100 U/ml penicillin, 100 µg/ml streptomycin and 10 % fetal bovine serum (37 °C, 5 % CO_2). Medium was replenished or refreshed totally every other day. Differentiation was induced by addition of 1 µM RA.

Raji cells (suspension) were cultured in completed DMEM consisting of DMEM medium supplemented with 2 mM L-glutamine, 100 U/ml penicillin, 100 µg/ml streptomycin and 10 % fetal bovine serum (37 °C, 5 % CO_2). Medium was replenished or refreshed totally every other day.

Cell cycle synchronization

Raji and THP-I cells were synchronized by means of a double thymidine block. A first blocking was reached by exposure to an excessive amount of thymidine (2 mM) overnight. After this incubation period the cells were released again for 7 to 8 hours by washing off the thymidine. Subsequently a second overnight thymidine block was applied, after which the cells were synchronized at the early S phase.

Propidium iodide staining

The results of the cell cycle synchronization can be monitored with flow cytometry. To this end, the cells were stained with propidium iodide (PI) (ex. 488 nm, em. 617 nm). Prior to staining, the cells were fixated in ethanol at 4 °C during at least 2 hours. After washing the cells with phosphate buffered saline (PBS), staining was performed in the dark during 30 min (50 µg/ml PI and 10 µg/ml RNase in PBS).

Flow cytometry

Oct4-eGFP or PI levels were monitored after cell isolation and resuspension in flow buffer (1 % BSA, 0.1 % NaN_3 in PBS) with an FC500 (Beckman Coulter) using the CXP analysis software.

Fluorescence microscopy

Fluorescence microscopy images of eGFP expression (ex. 485 nm, em. 515 nm, exposure time 5000 ms) were acquired on an Axiovert 200M inverted fluorescence microscope equipped with the Axiovision multichannel fluorescence module and an AxioCam MRM camera (Carl Zeiss). Colonies (without auto-fluorescent medium) were screened at 10x magnification using a Carl Zeiss short distance Plan-Acromat objective and visualized using Zeiss filter set n° 38 (BP 470/40, FT 495, BP 525/50). For larger colonies, several images were stitched by use of Photoshop CS4 on TIFF images (Adobe).

Acid histone extraction

Cells were isolated using 0.25 % trypsin-EDTA. After a washing step, cells were resuspended in Triton extraction buffer (TEB; 0.5 % Triton X 100 and 0.02 % NaN_3 in PBS) at a cell density of $1x10^7$ cells/ml, and incubated for 10 min on ice. Subsequently the cells were centrifuged at 1500 rpm for 10 min at 4 °C, and resuspended in half the volume of TEB. After centrifuging again, the cells were resuspended in 0.2 N HCl at a cell density of $4x10^7$ cells/ml and incubated overnight at 4 °C, after which the supernatant containing the histones, was isolated. Protease inhibitors (Roche, 11836170001) were freshly added to each buffer for every experiment, unless stated otherwise. Protein concentration of the extracts was measured by Bradford/Coomassie assay.

Direct boiling in Laemmli buffer

Direct boiling in SDS-loading dye for protein isolation was performed. Human ESC were directly boiled (3 min at 100 °C) after harvesting with trypsin, in 2X Laemmli buffer (4 % SDS, 20 % glycerol and 10 % β-mercaptoethanol in 50 mM Tris (pH 6.8)).

Gel electrophoresis and Western blotting

Per sample the required amount of histone extract was vacuum-dried and dissolved in 2X Laemmli buffer, separated on a 15 % Tris-HCl gel (BioRad Laboratories) and tank-blotted on a nitrocellulose or a PVDF membrane. The membrane was incubated overnight at room temperature with the appropriate antibody. The following antibodies were purchased from commercial vendors: a C-terminus directed histone H3 Ab (1:1000; Abcam, ab10799), an N-terminus directed histone H3 Ab (1:1000; Merck Millipore, 05-499) and an antibody specific for the H3 N-terminus when cleaved after A21 (H3.cs1 Ab; 1:1000; Active Motif, 39573) anti-cathepsin L catalytic domain antibody (1:1000; Abcam, ab49984), anti-

cathepsin L mature domain antibody (1:500; Abcam, ab58991), anti-cathepsin V antibody (1:1000; Abcam, ab49982), anti-neutrophil elastase antibody (1:500; Abcam, ab21595). All blots were detected by chemiluminescence using a Versadoc imaging system (BioRad Laboratories). Biotinylated histone H3 (as described below) was specifically visualized using HRP-conjugated avidin (45 min incubation; 1:100000; eBioscience, 18-4100-94). When needed, blots were stripped by incubating at 50 °C for 4 to 6 hours in stripping buffer containing 2 % SDS, 0.1 M β-mercaptoethanol, 0.05 M Tris (pH 6.8).

In-gel digest of propionylated histones

Vacuum-dried histones were propionylated as described [197] and separated by gel electrophoresis. Sypro Ruby (Invitrogen) stained gel pieces were cut out and subsequent in-gel digestion was performed as described before [198], with only a slight modification to the protocol: 25 mM ammonium bicarbonate (ABC) buffer was used and the alkylating agent applied was iodoacetamide (100 mM). After peptide extraction out of the gel pieces, samples were vacuum-dried and a second round of propionylation was completed to propionylate the newly generated N-termini.

Reverse transcription – quantitative PCR

RNA isolation, cDNA preparation and reverse transcription – quantitative PCR (RT-qPCR) analysis were performed as described before [199]. In short, cells were suspended in Trizol after isolation and stored at -80 °C. After reverse transcription (SuperScript II kit, Invitrogen), qPCR analysis was performed for pluripotency gene *POU5F1* (Taqman assay Hs01895061_u1) with normalization to an optimized pool of 3 reference loci, namely *B2M* (RTPrimerDB ID #2), *RPL13A* (#6) and AluSq (Forward: CATGGTGAAACCCCGTCTCTA – Reverse: GCCTCAGCCTCCCGAGTAG) (all SYBR Green assays). The additional TaqMan assays used were: cathepsin L (Hs00377632_m1), cathepsin V (H100426731_m1), neutrophil elastase (Hs00975992_g1 and Hs00975994_g1), all purchased from Life Technologies.

Histone biotinylation

One mg of purified histone H3, isolated from calf thymus (Roche), was biotinylated using the EZ-Link Sulfo-NHS-Biotinylation kit (Thermo Scientific, 21425), according to the manufacturer's conditions. After the biotinylation reaction, the excess of biotin-label was removed by addition of 5 % hydroxylamine (HA)-solution in a concentration of 6 µl HA per 100 µl sample.

N-terminal sequence analysis

A histone extract taken one day after induction of differentiation and known to contain clipped H3 (cH3) as validated by Western blotting, was blotted onto PVDF and proteins were visualized using Ponceau S staining. Subsequently the cleaved area was cut and subjected to Edman degradation, which was

outsourced to Eurosequence (Groningen, The Netherlands) and performed according to standard procedures.

MS method

Propionylated peptides were dissolved in 0.1 % formic acid in water (buffer A) and separated on a PepMap 100 (C18) column (I.D. 75 µm, length 25 cm, particle size 5 µm) by use of a U3000 LC-system (Dionex) at a flow rate of 300 nl/min. Elution was performed with 80 % acetonitrile/0.1 % formic acid (buffer B) using a gradient of 10 % to 60 % buffer B in 60 min. A Q-TOF Premier mass spectrometer (Waters) with nano-ESI source was operated in the data-dependent mode, with a resolution of 10.000. Survey MS scans were acquired (*m/z* 425-1300) and up to 7 precursors (*m/z* 50-2300) with charge state 2+, 3+ or 4+ exceeding the signal threshold were isolated for fragmentation by collision induced dissociation, using the collision energy profile as suggested by the manufacturer. An inclusion list contained precursor *m/z* value 574.3 since preliminary experiments suggested that this is a cleavage site fragment.

MS data analysis

Database searching was performed against a custom-made database containing human histone sequences obtained from the National Center for Biotechnology (NCBI) database, using a Mascot 2.3 in house server (Matrix Science). Mass error tolerances for the precursor ions and its fragment ions were set at 0.35 Da and 0.45 Da respectively. Enzyme semi-specificity was set to Arg-C, allowing for up to two missed cleavage sites. Variable modifications included acetylation, dimethylation and propionylation on lysine, methylation and dimethylation on arginine and oxidation of methionine. Lysine monomethylation was searched as the sum of propionylation and methylation since monomethylated lysine residues can still be propionylated. N-terminal propionylation and carbamidomethylation of cysteine residues were set as fixed modifications. Redundant peptides were filtered including only the highest scoring match under the highest scoring protein containing that match. Low confidence identifications were excluded using an expectancy value cut-off of 0.05. Nevertheless all of the used spectra were additionally manually validated using Mascot Distiller software (Matrix Science).

Nuclear extraction and protease inhibitor experiments

Nuclear extracts (NE) were prepared using the commercially available Episeeker Nuclear Extraction Kit (Abcam, ab113474), omitting protease inhibitors during extraction. Calf H3 (1 µg) and NE (0.5 µg) were incubated together in buffer (150 mM NaCl, 2 mM β-mercaptoethanol, 0.1 mM EDTA, 10 % glycerol and 25 mM Tris-HCl (pH 7.5)) at 37 °C for 1 hour, after which the assay was inactivated by incubation at 99 °C for 10 min. All samples were analyzed by Western blotting.

The protocol applied for isolating the nuclear fraction of THP-I cells used in the first step a hypotonic lysis buffer (10 mM Tris (pH 8), 1 mM KCl, 1.5 mM $MgCl_2$, 1 mM dithiothreitol, PIC), in which the cells were incubated for 30 min at 4 °C. After centrifugation (10 min, 4 °C, 2000 rpm), the isolated nuclei were resuspended in R1-buffer (40 mM Tris, supplemented with tributylphosphine, phosphatase inhibitors, PIC and endonuclease) and sonicated for 10 min.

For the inhibition assay the following commercially available inhibitors were used (Figure 4.11): 4-(2-Aminoethyl)-benzenesulfonyl fluoride hydrochloride (AEBSF, Sigma-Aldrich, 76307), E64 (Sigma-Aldrich, E3132), EDTA (Sigma-Aldrich, E5134), Bestatin (Sigma-Aldrich, B8385), Pepstatin A (Sigma-Aldrich, P5318), protease inhibitor cocktail tablets (PIC, Roche, 11836170001) and three specific cathepsin L inhibitors (I (219421), III (219427) and CAA0225 (219502), purchased from Calbiochem. Inhibitors were pre-incubated with the NE for 15 min at 37 °C prior to addition of the H3 substrate. Their inhibitory capacity was validated and concentrations yielding aspecific inhibition were excluded from the assay.

In vitro enzymatic incubation assays

Recombinant active cathepsin L was purchased from Abcam (ab81780). Recombinant pro-cathepsin V was obtained from Enzo Life Sciences (BML-SE554-0010) and could be activated by pre-incubating the proenzyme in assay buffer for 5 to 30 min. The pH 5.5 buffer consisted of 10 mM HEPES, 10 mM KCl, 1.5 mM $MgCl_2$, 0.34 M sucrose, 10 % glycerol and 5 mM β-mercaptoethanol. The pH 7.5 buffer was composed of 25 mM Tris-HCl, 150 mM NaCl, 0.1 mM EDTA, 10 % glycerol and 2 mM β-mercaptoethanol. Incubations were performed at 37 °C for 1 hour.

4.4. Results

4.4.1. Feeder-Free cultured hESC show continuous histone H3 cleavage during differentiation

To verify whether the histone H3 N-tail is cleaved during hESC differentiation in analogy to mESC [184], two hESC lines were monitored: an Oct4-eGFP reporter hESC cell line that expresses eGFP under the control of the Oct4 promoter and a non-reporter cell line UGENT2. Both cell lines were cultured in feeder-free conditions during 5 days after induction of differentiation by addition of 2 μM RA to the culture medium, in the absence of bFGF. Cells were isolated every 24h and histone extracts were prepared.

The differentiation status during the experiments was validated for both cell lines. Morphologically each cell line displayed clear traits of differentiation, resulting in more lengthened cells and loss of round colony shape (data not shown). To further validate the efficiency of differentiation, Oct4-eGFP reporter cells were directly monitored by both flow cytometry and fluorescence microscopy [200]. Flow data showed a clear drop in eGFP signal through time (Figure 4.1A), and microscopy imaging visualized a sustained but definite reduction in eGFP throughout the hESC colonies (Figure 4.1B). To monitor loss of pluripotency in the conventional non-reporter stem cell line (UGENT2), RT-qPCR analysis of the pluripotency gene *POU5F1* encoding for the transcription factor Oct4, coordinately confirmed differentiation in these cells (Figure 4.1C).

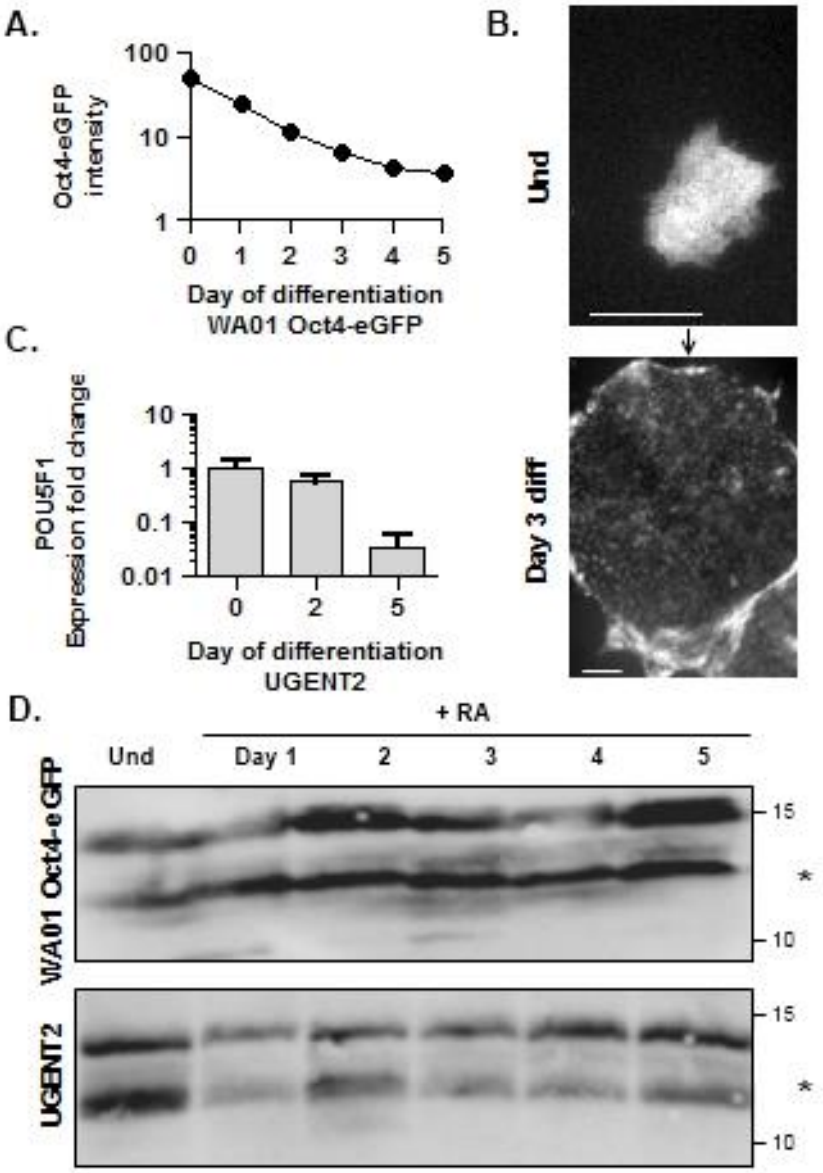


Figure 4.1 Histone H3 cleavage in hESC under feeder-free culture conditions.
Human ESC samples were collected during five days after differentiation induction with RA. Differentiation was confirmed by **A**: flow cytometry Oct4-eGFP analysis for the Oct4-eGFP cell line. In addition, **B**: fluorescence microscopy data affirmed a drastic drop in Oct4-eGFP signal when comparing an undifferentiated colony (Und) and a colony after three days of differentiation (scale bars represent 200 µm). Also **C**: RT-qPCR analysis illustrated the decrease in POU5F1 expression (encoding for Oct4) over time for the UGENT2 cell line (mean and standard deviation of 2 replicates). **D**: Western blot analysis (2 µg histone extract per sample) using a C-terminal H3 antibody reveals N-terminal H3 clipping (cleaved H3 indicated with an asterisk) in both the Oct4-eGFP reporter (upper image) and UGENT2 cell line (lower image).

Samples were monitored for histone H3 cleavage by means of Western blotting analysis, using an H3 C-terminally directed antibody. For both cell lines, lower molecular weight bands of histone H3 (indicated with an asterisk) were visualized at each time point throughout the differentiation experiment, indicating N-terminal cleavage of H3 (Figure 4.1D).

4.4.2. **Continuous histone H3 cleavage in the Oct4-eGFP reporter hESC line is not related to the Oct4 expression level**

The results above differ substantially from the data reported for differentiating mESC, which display a pattern with upcoming cH3, reaching a maximum intensity after 4 days, and decreasing again towards the end of the experiment [184]. The continuous H3 clipping found in feeder-free cultured hESC could be best explained by assuming that the timeframe was not extended enough to demonstrate the disappearance of cH3. The cH3 pattern might also be influenced by the speed and heterogeneity of differentiation, as also shown by Duncan et al., where spontaneous and RA induced differentiation and EB formation all lead to different cleavage patterns in terms of timescale. Therefore we repeated the experiment with the Oct4-eGFP reporter cell line, monitoring the hESC in feeder-free culture for a longer period of time (14 days), and subjected them to three different methods of differentiation to vary the speed at which they lose stemness throughout this timeframe: **(i)** in their undifferentiated status by culturing the cells in E8 medium, **(ii)** inducing spontaneous differentiation by omitting bFGF from the culture medium and **(iii)** enforcing directed differentiation by addition of RA. Morphological changes were observed using light microscopy (data not shown), global Oct4 expression and differentiation were monitored using flow cytometry. According to the flow plots shown in Figure 4.2A, the three used culture conditions indeed differ greatly in their (non)maintenance of stemness throughout the experiment. When analyzing the hESC, it is clear that E8 medium promoted the undifferentiated status in contrast to both differentiation methods which led to a decreased Oct4 expression. While spontaneous differentiation divided the culture system into two groups with loss of the Gaussian distribution in the flow histogram, RA induced a collective decrease in Oct4 expression of all cells resulting in a preserved Gaussian distribution shifting towards lower fluorescence. Also, differentiation induced by RA occurred much faster and a minimal Oct4 expression was reached after 12 days of differentiation while spontaneous differentiation continued at a slower pace and did not reach a minimum limit in Oct4 expression after 14 days of differentiation.

Despite the clear differences in Oct4 expression as described above, all time points in each culture condition appeared to display clear N-terminal H3 clipping when monitored with Western blotting (Figure 4.2B).

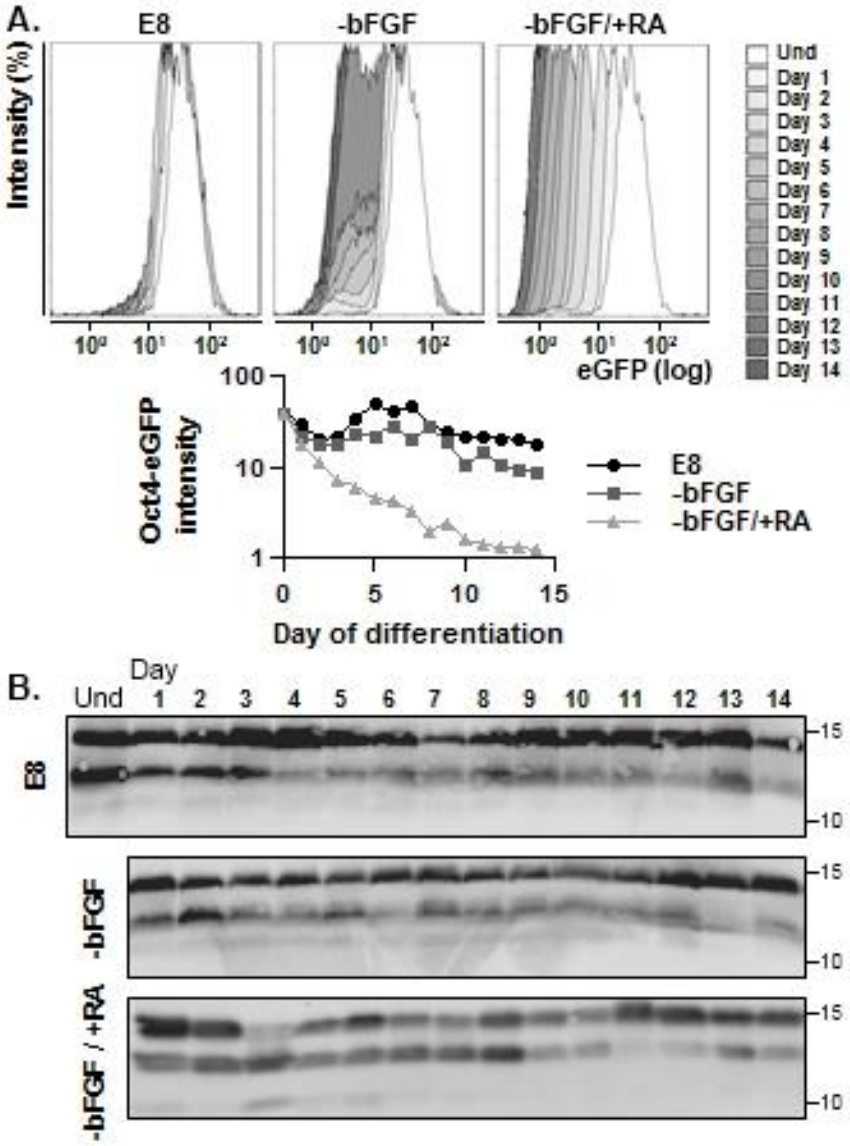


Figure 4.2 H3 cleavage in feeder-free cultured hESC under application of different culture media within an extended time frame.

The Oct4-eGFP reporter cell line was cultured for 14 days in different conditions: **(i)** maintaining the pluripotent state in E8 medium, **(ii)** differentiating spontaneously (in the absence of bFGF) and **(iii)** differentiating after RA induction. **A**: Flow cytometry analysis displays a definite reduction in Oct4-eGFP levels under differentiation-allowing circumstances. The flow histogram plots show a variable shift in Oct4-eGFP signal under differentiating conditions: whilst the signal in RA-stimulated cells decreases collectively for the total population, the spontaneously differentiating cells are divided into two groups with different paces of differentiation. All data shown on the flow histogram depict data from the undifferentiated stage until day 14 of differentiation, with the most anterior graph representing the Oct4-eGFP signal from undifferentiated cells, and the graph most at the back illustrating day 14. **B**: Western blot images (2 µg samples, C-terminal H3 antibody) show continuous cH3 formation in hESC, regardless of the hESC culture condition.

4.4.3. **Histone H3 clipping in hESC is not an *in vitro* artifact**

To exclude the possibility that this continuous histone H3 cleavage in a feeder-free culturing system is actually an *in vitro* artifact induced during extraction, we spiked biotinylated calf histone H3 during hESC

histone extraction at day 1 of differentiation for subsequent avidin-HRP detection on Western blot. We initially confirmed that indeed the biotinylated histone H3 still is susceptible to proteolytical degradation (Figure 4.3A).

After spiking biotinylated calf H3 during extraction, a lower band was detected on the Western blot from the histone extracts only when using the C-terminal H3 antibody. This band was absent when immunoblotting was done using avidin-HRP, which specifically highlights the biotinylated fraction (Figure 4.3B). This implies that the histone H3 cleavage was indeed already present during culturing and that the truncated form was not created during extraction. Moreover, when the extraction was performed both with and without protease inhibitor cocktail (PIC), no difference between these samples was found, confirming that no additional cleavage was induced during extraction. Although we verified the enzymatic susceptibility of biotinylated H3,we also repeated this experiment by using non biotinylated calf H3. Here, we added an excessive amount of calf H3 to the buffers during the extraction procedure. Since western blot analysis can not distinguish between this calf H3 in and endogenous H3, we compared intensities of the lower cH3 band between the samples with or without spike in (data not shown). The relative abundance of clipped H3 was clearly lower in the HE spiked with calf H3, suggesting no additional proteolysis *in vitro*. Although we cannot exclude that this was due to an inhibitory effect of the excess calf H3 towards the clipping of the endogenous H3, this result at least suggests that the added calf H3 was probably not clipped to a notable amount during extraction and did not contribute to the cH3 content in the hESC samples

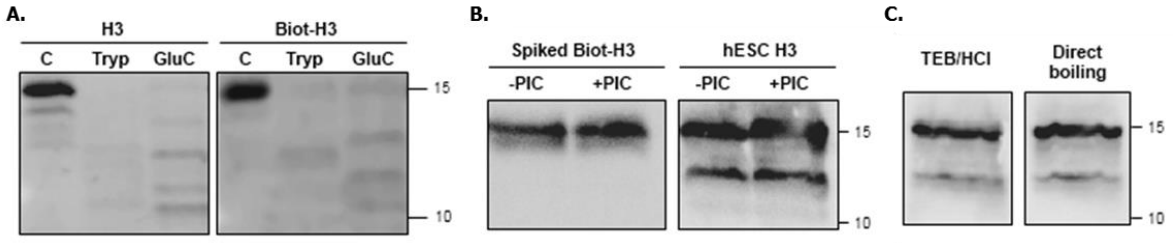


Figure 4.3 Histone H3 cleavage is not an in vitro artifact.
A: The digestion of purified calf H3 and biotinylated calf H3 with both trypsin or GluC was compared, and Sypro Ruby staining after SDS-PAGE illustrated the analogous susceptibility of both samples. **B**: Comparison of spiked biotinylated H3 (Biot-H3) with the H3 content extracted from hESC shows that only the latter undergoes truncation. As both immunoblotting images demonstrate, including the PIC (-PIC versus +PIC) during the extraction procedure does not influence the outcome. Per lane 2 µg sample was loaded. **C**: Illustrates the comparison of two different extraction protocols: TEB/HCl extraction (left, 2 µg sample) versus direct boiling in Laemmli buffer (right, 5 µl loaded). Western blotting analysis shows no difference in cH3 content between those protocols.

In addition, the TEB/HCl extraction protocol was compared with direct boiling of the hESC after harvesting. Directly boiling in Laemmli buffer diminishes the steps of the extraction procedure and thus the steps in which cH3 can be artificially formed. For this experimental set-up, Oct4-eGFP reporter hESC

were differentiated with 2 µM RA for 24h. After this the hESC were harvested and split into two to carry out both protocols simultaneously on the same starting material. As can be seen in Figure 4.3C, both protocols result in the same Western blotting image containing cH3.

Finally, besides the experiments to exclude possible unwanted *in vitro* clipping activity during the histone extraction protocol, we also briefly looked into whether or not the other core histones are processed during our hESC experiments. In the MS data obtained in search for the clipping sites (Results 4.4.4), we could not find any evidence for other clipped histone forms. Additionally, we analyzed several samples from differentiating hESC by means of Sypro Ruby staining after SDS-PAGE to visualize possible additional protein bands indicative of other degradation products. Again, no obvious changes in histone patterns were found, except for that of histone H3 (Figure 4.4). Although this is evidently not stringent enough to affirm their absence, it does add to the evidence for the specificity of this clipping event.

Taking all these data together, we conclude that the detected H3 fragments must have been generated prior to the extraction procedure in our experiments. Nevertheless, undisputable confirmation hereof could only be attained by detecting the specific fragment in intact cells, e.g. by live-cell immunofluorescence staining with the H3.cs1 antibody. Yet, we consider the evidence outlined here adequate to imply a biological origin of the cH3 fragments in differentiating hESC.

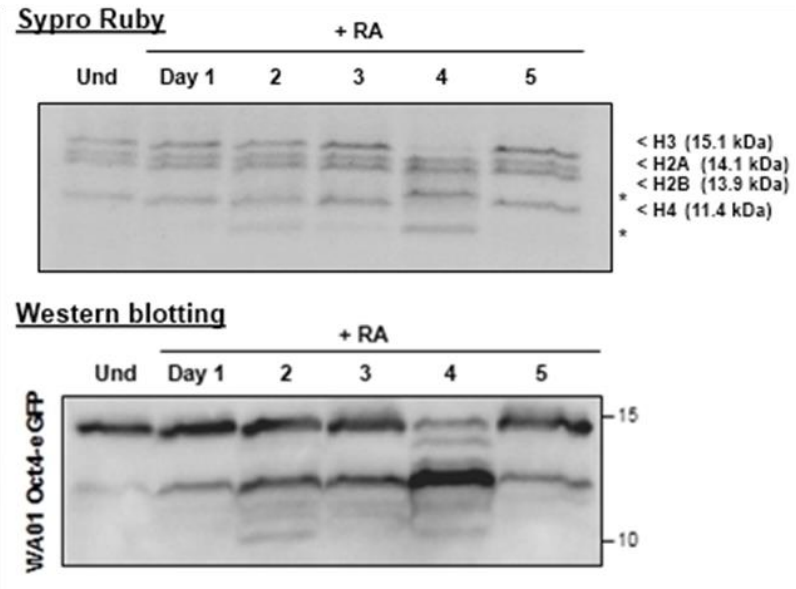


Figure 4.4 No general histone clipping in hESC.
Sypro Ruby staining after SDS-PAGE (upper panel) shows no notable changes in any other core histone content other than histone H3. Specific western blotting analysis (lower panel) with the C-terminal H3 antibody clearly showed the cH3 content. Samples were obtained from a differentiation (2 µM RA) experiment performed on Oct4-eGFP reporter cells (MEF culture) during 5 days. The input for each lane was 3 µg histone extract.

4.4.4. **Similar histone H3 cleavage sites are found in mouse and human ESC**

To verify if this H3 truncation event has any parallel to the clipping event reported in mESC, we next set out to identify the cleavage site(s).

Applying the antibody directed against the C-terminus of H3, cH3 is visualized, in some cases as multiple bands. As opposed to the C-terminal antibody, immunoblotting with an antibody directed against the N-terminal end of histone H3 detected no cH3 for both the Oct4-eGFP reporter and UGENT2 cell line (Figure 4.5A and B). Only the band of intact histone H3 was seen, indicating N-terminal cleavage. When using the C-terminal antibody, the distance of the most intense cH3 band to the intact H3 form, implies a loss of approximately 3 kDa, which roughly corresponds to about 30 Aa. Using the histone H3.cs1 antibody that was developed to specifically detect H3 truncated after A21 [184], a weak signal was detected for samples of the Oct4-eGFP cell line, suggesting A21 as a possible target site also in hESC.

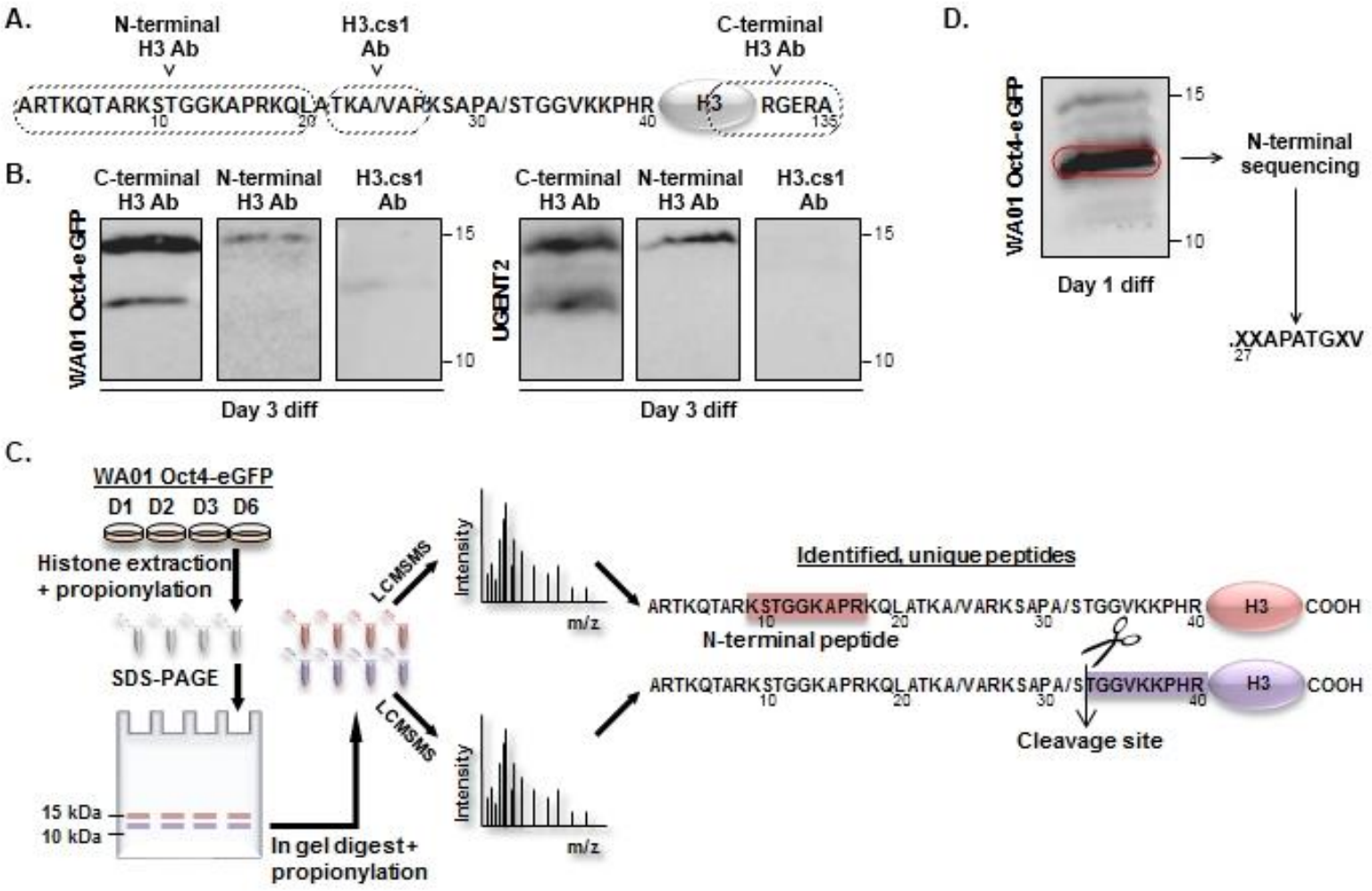


Figure 4.5 H3 cleavage in feeder-free cultured hESC under application of different culture media within an extended time frame.
The application of differentially directed antibodies for immunoblotting (antibody epitopes displayed in **A)** illustrates that the cleavage is situated N-terminally, and proposes alanine 21 as one of the possible cleavage sites **(B). C:** MS analysis after propionylation of different hESC samples (Day 1, day 2, day 3 and day 6 of differentiation) also reveals residue 31 as a site of cleavage. **D**: An additional site, arginine 26, was annotated by Edman degradation.

Apart from the most prominent A21, Duncan et al. found additional histone H3 cleavage sites in mESC, so we pursued further investigation of the presence of any other possible cleavage site(s) in hESC. A specific MS approach for histone analysis was therefore optimized which is subsequently described in more detail. A regular bottom-up approach using trypsin would cleave after each lysine (K) or arginine (R) (except when followed by a proline), resulting in unidentifiably small peptides because of the abundant presence of these basic Aa in histones. We thus propionylated the histones prior to digestion, which modifies all free primary amine groups (the N-termini and the ε-amino group of unmodified and monomethylated K). Of note, when the N-terminus or a lysine residue is covered by an endogenous modification other than monomethylation, this residue cannot be propionylated but is also blocked from proteolytical digestion by that modification itself. Consequently, trypsin now only cleaves C-terminal to arginine residues mimicking an Arg-C digestion for histone H3, resulting in larger, identifiable peptides [128]. By post-digestion propionylation all newly generated N-termini are subsequently being propionylated as well. The specific cleavage sites present in cultured hESC can then be detected as peptides containing a non-arginine C-terminus or which start N-terminally with an Aa not neighboring an arginine. Spectra corresponding to such so-called semi-Arg-C peptides were additionally manually validated.

Hence we separated propionylated histone extractions of the feeder-free cultured Oct4-eGFP reporter cell line at four different time points after RA induction (D1, D2, D3 and D6) on SDS-PAGE (Figure 4.5C). This way, multiple molecular weight bands could be cut out for subsequent in-gel digestion using trypsin. After a second round of propionylation, samples were analyzed by RP-LC and ESI-MSMS. Of all identified histone H3 peptides, 100 % ended C-terminally with R and 94.4 % had an N-terminus adjoining an R, confirming successful propionylation.

First, to confirm the N-terminal histone H3 cleavage by MS we monitored the presence or absence of an N-terminal histone H3 peptide in the intact H3 and the cH3 gel bands. The peptide $K_{[9]}$STGGKAPR starting after $R_{[8]}$ covers the N-terminus of histone H3 and was only identified in the H3 bands and not the cH3 (Figure 4.5C), in contrast to the globular histone H3 peptides and the C-terminal $V_{[117]}$TIMPKDIQLAR which were present in both. Two differentially modified semi-Arg-C peptides both started at residue 32 which N-terminally does not flank an R: $T_{[32]}GGVK_{prop}K_{prop}PHR$ and $T_{[32]}GGVK_{prop}K_{me\text{-}prop}PHR$ (Figure 4.6). Since these spectra correspond to the peptide generated by cleavage after Aa 31 and this residue itself is not represented in the spectrum, it is impossible to define whether Aa 31 is an alanine or serine, coming from the H3.1/H3.2 or the H3.3 isoform respectively. At each of the four different time points $A_{[31]}$ or $S_{[31]}$ clipping was identified in the gel pieces containing cH3, but not intact H3. Of note, using this MS method no cleavage at an arginine residue can be detected, since the protocol

itself introduces clipping after arginine during digestion. Indeed, by applying Edman degradation, we also assigned R26 as a cleavage site (Figure 4.5D). Since the Aa at position 31 is identified by Edman degradation as an alanine this cleaved form is derived from the H3.1 and/or H3.2 isoforms.

In conclusion, both Western blotting and MS confirmed N-terminal cleavage with an intact C-terminus, and the latter assigned Aa 31 as a cleavage site for histone H3 without isoform specification. N-terminal sequence analysis added R26 as a cleavage site, at least for the H3.1 and/or H3.2 isoforms.

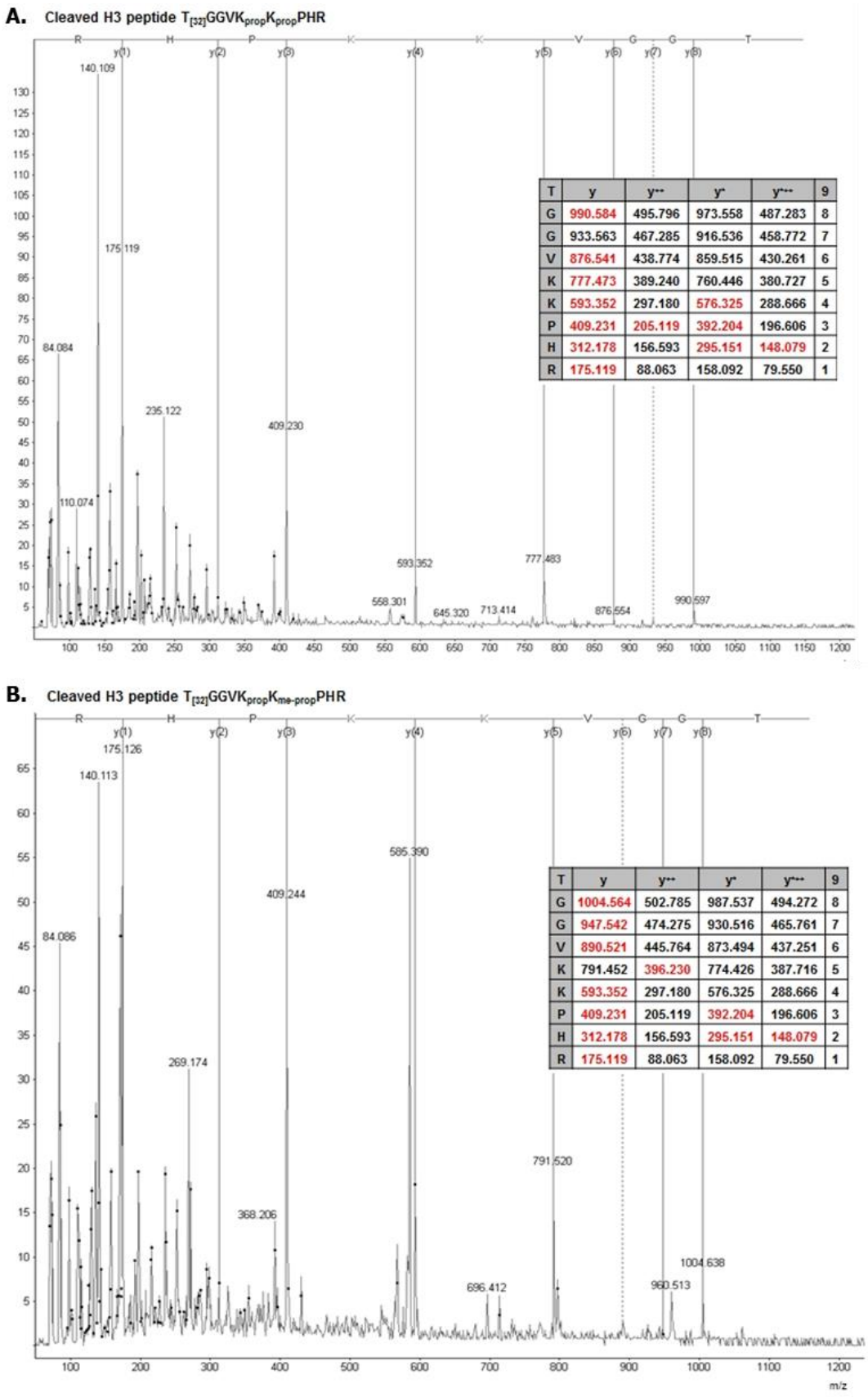


Figure 4.6 Manual validation of cleaved H3 peptides.
Validation of the MSMS spectrum representing **A**: T[32]GGVKpropKpropPHR and **B**: T[32]GGVKpropKme-propPHR respectively. Matching fragments are displayed in red.

4.4.5. When grown on MEF hESC cleave H3 in an identical temporal window compared to mESC

Since mESC differentiation showed various temporal clipping patterns depending on the differentiation protocol applied, we further extended the different culture conditions to also include the more classical culture system on a MEF feeder layer. Although these MEF are mitotically inactivated prior to cell culturing, cH3 formation derived from MEF cannot be excluded. Thus, RA was first added to a culture consisting only of a confluent layer of MEF, without hESC present. No cleavage was detected by immunoblotting (Figure 4.7).

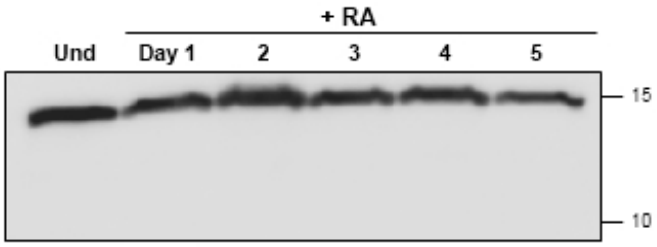


Figure 4.7 MEFs display no histone H3 cleavage.
When incubating a culture consisting only of MEF with RA for 5 days, no cH3 could be detected.

When RA was added to hESC cultured on MEF, both morphological assessment of the differentiated status, and flow cytometry as well as RT-qPCR confirmed a decrease in Oct4 level as differentiation proceeded for the Oct4-eGFP reporter and UGENT2 cell line respectively (Figure 4.8A and B). Of note, as seen by fluorescence microscopy within one colony patches or 'islets' of undifferentiated cells became apparent, indicating heterogeneous stemness within colonies (Figure 4.8C). Remarkably, this Oct4 expression pattern is clearly different from what was seen when growing hESC feeder-free, which showed a diffuse expression pattern in the abovementioned experiments.

Surprisingly, when analyzing these histone extracts with Western blotting, a more 'mouse-like' pattern of H3 cleavage becomes visible: on day 2 of differentiation cH3 appears, attains a maximum on day 4 and fades considerably on the last day of the experiment. This pattern was visualized for both cell lines, again confirming the cleavage capability of both the reporter and non-reporter cell line (Figure 4.8D).

Also here the N-terminal nature of the H3 cleavage was double-checked using the N-terminally directed antibody, which visualized only the band of intact histone H3. In addition, a clear signal was detected using the H3.cs1 antibody for the UGENT2 sample, once more pointing out A21 as a possible cleavage site (Figure 4.8E).

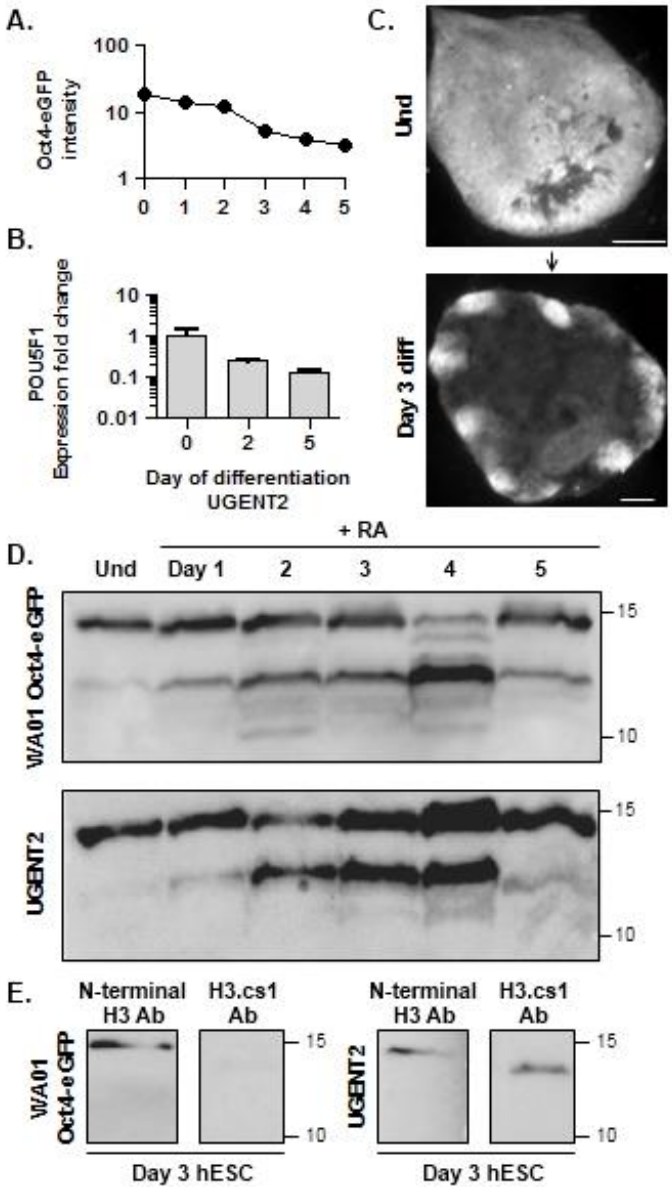


Figure 4.8 A different H3 cleavage pattern is seen in MEF cultured hESC.
Aside from the application of feeder-free culture, additional differentiation experiments with hESC cultured on a MEF feeder layer were conducted. The differentiated status is again confirmed by **A**: flow cytometry and **B**: RT-qPCR (mean and standard deviation of 2 replicates). **C**: Fluorescence microscopy analysis indicates differentiation heterogeneity among colonies, as within each colony several 'islets' of concentrated Oct4-eGFP expression were seen (scale bars represent 200 µm). **D**: Western blotting analysis reveals that when hESC are cultured on MEF, a cleavage pattern is shown that resembles more to the mESC results. **E**: Also, the N-terminal orientation of the cleavage and A21 as a possible cleavage site is confirmed by application of the N-terminal and H3.cs1 antibody respectively.

In order to better understand the biological turnover of this H3 clipping event in MEF-cultured hESC, the initial differentiation experiments were also repeated both in a more narrow time frame and over a more extended time period. For this experiment two cell lines were combined (UGENT1 and UGENT2). Figure 4.9 shows both UGENT1 and UGENT2 display a pattern of upcoming and decreasing H3 proteolysis. This pattern generated by harvesting cells every 4 hours, is even more undulating than the one obtained by

daily sampling of the same cell line (Figure 4.8) and is in stark contrast with the continuous cleavage found in feeder-free cultured hESC.

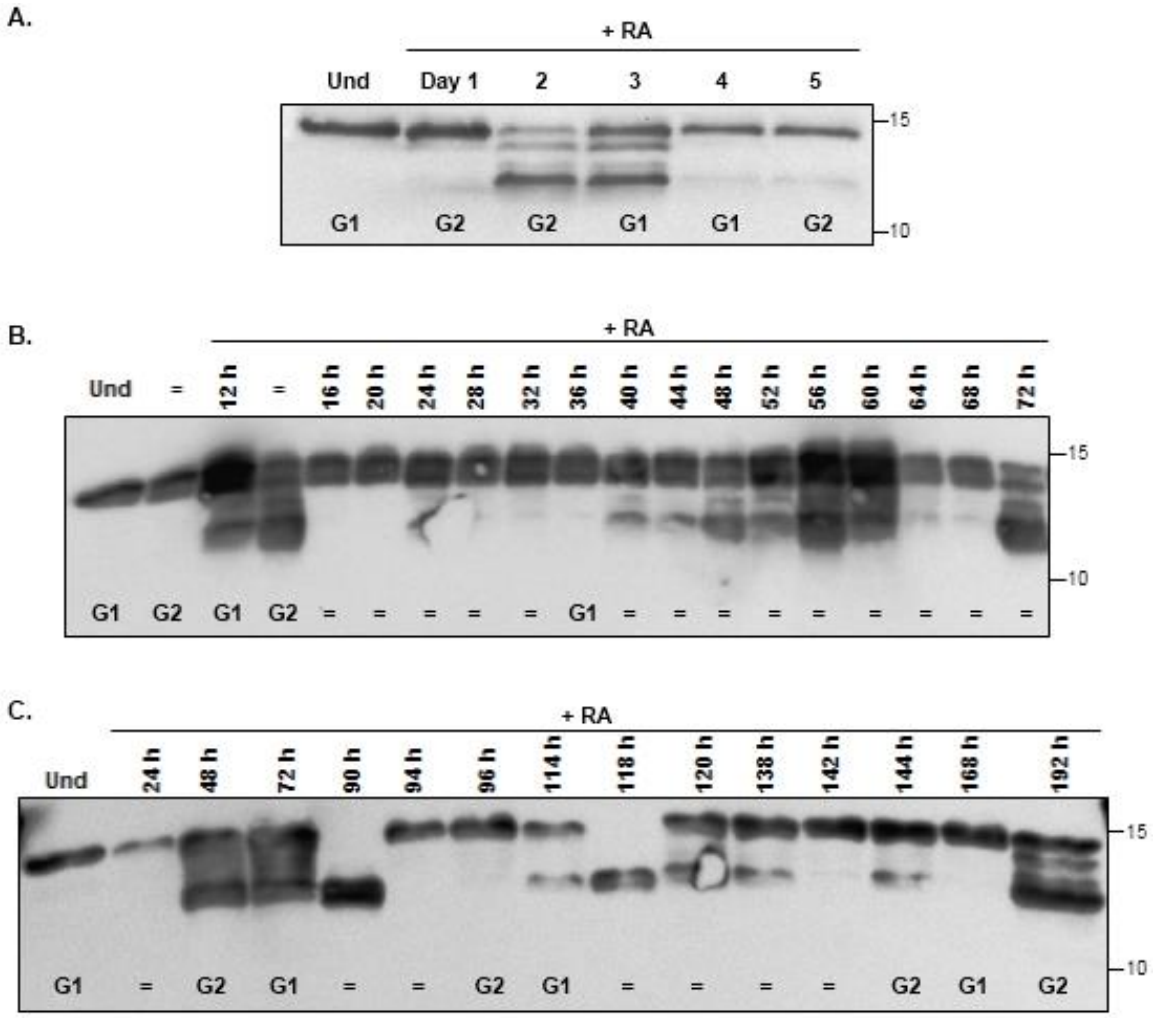


Figure 4.9 Undulating pattern of H3 clipping in MEF cultured hESC.
A: The initial differentiation experiment showed cleavage for both the G1 and the G2 cell line, with upcoming cH3 on day 2 and fading again on day 4 of differentiation. **B**: When increasing the time resolution on day 3, 4 and 5 of differentiation, western blotting analysis shows a 'wave' pattern of upcoming and decreasing clipping. **C:** The same result is obtained on a larger time scale when the experiment was extended up to 12 days. Also for B and C samples from both cell lines were used, as indicated on the blot image. Respectively 3 µg (B) and 2 µg (C) histone extract were loaded per lane. The time points of sample collection are indicated on all images ("x hours/days" after onset of differentiation).

4.4.6. **H3 proteolysis is not cell cycle related**

Both this undulating pattern (section 4.4.5) as well as the synchronizing effect of RA on the stem cell cycle directed us towards the assumption that this H3 clipping event might be cell cycle related. We hypothesized that the histone clipping event takes place during the S phase, in which the DNA is duplicated and the histone content is remodeled accordingly. As the chromatin structure is loosened upon replication, it becomes more susceptible for proteases, which might fulfil specific as well as

aspecific enzymatic activity. To investigate this we performed a cell synchronization experiment with a Burkitt's lymphoma Raji cell line. Treatment with double thymidine results in a G1/S-phase arrested cell population. The use of PI staining in combination with flow cytometry allows progression of the cells through the cell cycle to be monitored (Figure 4.10 A). The terminally differentiated raji cells showed no clipping activity, disregarding whether they are synchronized or not (Figure 4.10 B). Of note, the progenitor THP-I cell line (acute monocytic leukemia) is capable of H3 proteolysis. However, this clipping potential seems to be unrelated to differentiation, as the addition of RA did not induce any changes. These results are in accordance with the results in section 4.4.2 that indicate Oct4 expression, and therfore differentiation, does not seem to be a major requirement for histone clipping. Also here, cell cycle synchronization did not alter the constant pattern of clipping, even though the obtained THP-I results were not as consistent as other cell lines.

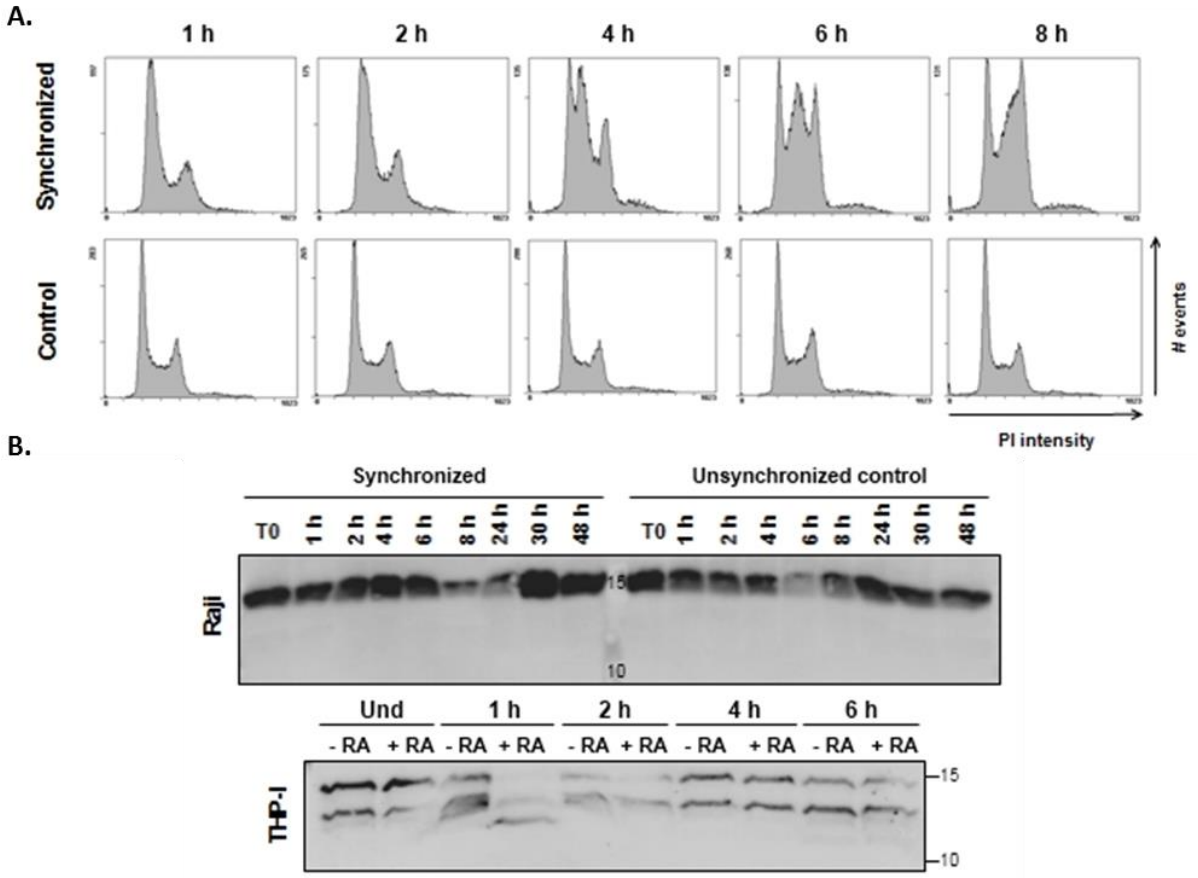


Figure 4.10 Histone H3 clipping in synchronized Raji cells and an (un)differentiating progenitor THP-I cell line

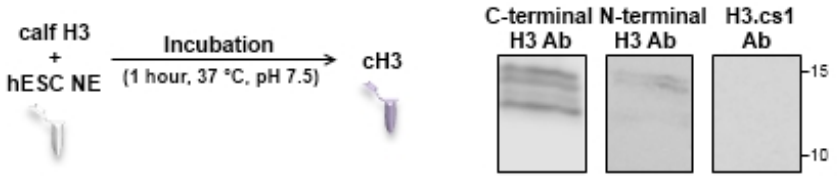
A: Validation of cell cycle synchronization. Flow plots after PI staining show a clear shift in population distribution after release from the thymidine block, compared to the steady profile of the control samples. **B**: No cH3 was observed in Raji cells in western blotting analysis, neither in unsynchronized or synchronized conditions (T0 = starting point right after release of second thymidine block, other time points indicate number of hours after release). THP-I cells on the other hand are capable of H3 clipping, but the addition of RA does not influence the clipping pattern (Time points given are the number of hours after onset of differentiation).

4.4.7. **Histone H3 clipping activity in hESC is exerted by a serine protease**

In order to identify the clipping enzyme, we conducted different calf H3 incubation assays, whether or not with the application of various protease inhibitors. To this end, we prepared a nuclear extract (NE) from Oct4-eGFP hESC for which histone H3 cleavage was confirmed in the corresponding histone extract, and validated it to be capable of calf H3 proteolysis. Endogenous H3 was not detectable in the amount of NE used in these experiments.

In a first experiment we compared the incubation of H3 with hESC NE after immunoblotting with the three different antibodies described before. This confirmed that the cleavage pattern induced by the NE is similar to what was seen in the histone extract of Oct4-eGFP hESC: the cleavage is exclusively of N-terminal nature and there is no detection of A21 as a possible cleavage site (Figure 4.11A).

A.



B.

Inhibitor	Target	Concentrations applied		Inhibition
		Min (*)	Max (**)	
AEBSF	Serine proteases	0.1 mM	1 mM	+
E64	Cysteine proteases	1 µM	10 µM	-
PIC	Serine and cysteine proteases	1 tablet/10 ml buffer	1 tablet/2 ml buffer	+
EDTA	Metalloproteases	1 µM	10 µM	-
Bestatin	Aminopeptidases	1 µM	10 µM	-
Pepstatin A	Aspartyl proteases	5 µM	10 µM	-
Cath L inh I	Cathepsin L	1 µM	10 µM	-
Cath L inh III	Cathepsin L	10 µM	30 µM	-
Cath L inh CAA0225	Cathepsin L	5 µM	20 µM	-

C.

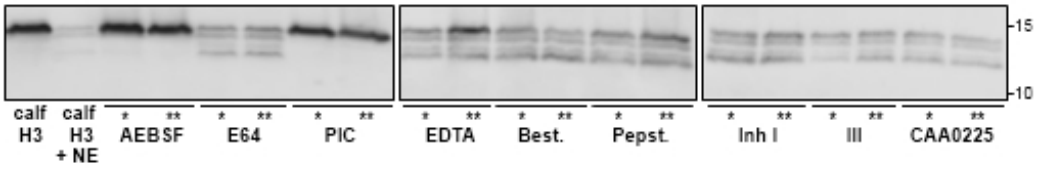


Figure 4.11 The histone H3 clipping activity in hESC is exerted by a serine protease.
A: In order to investigate the responsible cleavage enzyme, incubation assays were set up in which 1 µg calf histone H3 was incubated with 0.5 µg hESC NE. The NE was prepared from the Oct4-eGFP cell line and contains N-terminal histone H3 cleavage activity, as was confirmed in the corresponding histone extracts (data not shown). Immunoblotting analysis of the incubation product displays several cH3 fragments, with confirmed N-terminal origin. A21 could not be confirmed as a cleavage site (H3.cs1 Ab). The list of all inhibitors included, with their targets, the applied concentrations and their outcome in the inhibition assay is displayed in panel **B. C**: Western blotting analysis points out that only when using a serine protease inhibitor the cleavage activity of the NE can be inhibited. The latter is confirmed by the additional inhibitory effect of the PIC, targeting mainly serine and cysteine proteases.

In the following inhibition assay, several inhibitors were included (Figure 4.11B), comprising both general inhibitors for a certain protease enzyme class and specific inhibitors for cathepsin L, in parallel to the

mESC results, where cathepsin L was identified as the responsible clipping enzyme. Western blotting analysis pointed out that only AEBSF, a serine protease inhibitor, could establish inhibition of the cleavage, whereas all other inhibitors could not demonstrate any notable effect on the incubation. Also the PIC was able to inhibit the proteolytic activity on histone H3. The specific inhibitors for cathepsin L were not able to diminish the cleavage (Figure 4.11C).

4.5. Discussion

Histone clipping was reported for the first time even before these proteins received their current nomenclature [201,202]. Surprisingly however, these truncation events are still greatly understudied. Yet, from a practical point of view, techniques such as chromatin-immunoprecipitation run the risk of not detecting the substantial influence that histone clipping might have on their outcome. All techniques employing specific antibodies to the H3 N-terminus in general or to modifications thereof in particular, will fail in case of H3 proteolytic cleavage. Also MS-based (quantitative) comparisons of e.g. H3K4 and H3K27 methylations devaluate when dealing with samples where clipped histones are present.

Histone clipping has been reported in several distinct biological systems [187–190,192–195], entailing different biological settings and thus potential roles of this PTM. However, only recently it was admitted into the epigenetic landscape with the discovery of transient histone H3 clipping in differentiating mESC [184] and sporulating yeast [191]. The evolutionarily conserved sequence of histone proteins has been suggested to underlie the surprising occurrence of A21 clipping in both mouse and yeast, and we thus set out to verify whether the H3 clipping that accompanies mESC differentiation also takes place during hESC differentiation.

Duncan and colleagues monitored cH3 levels under several differentiation protocols in mESC, namely spontaneous monolayer differentiation after withdrawal of leukemia inhibitory factor (LIF), induced differentiation with RA and EB formation. Depending on the protocol applied, a different cH3 pattern was visualized. Where RA induction leads to an undulating pattern of upcoming and decreasing H3 clipping centered around day 2 and 3 of differentiation, EB formation displayed a faster migrating H3 band which peaked between day 8 and 12 but did not disappear completely after 14 days.

Here, we report that indeed histone H3 N-terminal clipping occurs in differentiating hESC and that its temporal appearance is equally influenced by the culture conditions. In feeder-free conditions, the two hESC lines tested (UGENT2 and WA01 Oct4-eGFP) show a continuous cH3 pattern after RA induced differentiation, resembling the results of EB formation in mESC. In contrast, when switching to culturing hESC on a feeder layer of MEF, both these cell lines obtain a pattern of upcoming cleavage appearing on day 1, reaching a maximum intensity at day 4 and fading again at the last day of the experiment,

similar to the temporal pattern described for RA induced mESC differentiation. Together, this suggests that within both human and mouse, the specific control over histone H3 cleavage during differentiation is profoundly influenced by experimental culture conditions applied. Although the cleavage event in mESC and hESC is appreciably similar, caution should be taken when functionally comparing these events, not in the least because hESC resemble more mouse epiblast stem cells than regular mESC. The latter are considered to represent a more homogeneous and naïve pluripotent state compared to hESC, which are designated to be in a heterogeneous and primed pluripotent condition and thus probably reflect a more developed state [21,24,203]. In contrast to Duncan et al., we could not find any clear correlation between the progression of differentiation and histone H3 clipping. The pluripotency status was monitored in this report by following Oct4 levels, i.e. specifically with an Oct4-eGFP reporter cell line. When monitoring three different differentiation methodologies, H3 cleavage seemed to be unaffected, while the Oct4 status was found to be influenced in the predicted way: a decrease in expression following (induced) differentiation. Considering the open and highly dynamic state of pluripotent hESC chromatin, this might not be such a surprise [55]. In line with this, on a Sypro Ruby stained SDS-PAGE gel we also observed some reduction in H4 band intensity at some time points (data not shown). Whether this indeed is a clipping event remains to be determined. Of note, when Duncan et al. inhibited cathepsin L, H3 clipping was abrogated while Oct4 expression still decreased, adding yet another argument against a direct link between Oct4 expression levels and histone clipping.

Histone H3 cleavage is not *in vitro* generated during the experiment as visualized here by the use of biotinylated and non biotinylated histone H3 during the extraction procedure. Also, protease inhibitors did not seem to influence the cH3 intensity, whether or not they are present. Moreover, both TEB/HCl extraction and direct boiling in SDS-PAGE Laemmli yields the same results, further supporting the endogenous formation of cH3. The fact that MEF themselves do not display any clipping, but can completely change the temporal appearance of cH3 in differentiating hESC, further argues in favor of a biologically regulated process. We also verified the occurrence of H3 clipping in a human, terminally differentiated cell line (Burkitt's lymphoma Raji cells) and these cells did not form any clipped histone H3.

If not generated *in vitro*, the link between cell cycle and this clipping event was investigated. First, synchronization was induced in terminally differentiated Raji cells by means of a double thymidine block, but no cH3 was seen. The cell cycle as such is thus not responsible for H3 proteolysis. Moreover, also the progenitor THP-I cell line was included and clipping was monitored after stimulation with RA and when subjected to cell cycle synchronization. Even though these results were not as consisitent as other cell lines, they show that THP-I cells are capable of H3 cleavage, with or without RA induction. However,

cell cycle synchronization did not alter the constant pattern of proteolysis found in these cells. Taken together, these findings suggest that clipping is not directly related to cell cycle, yet some degree of differentiation capacity is required for it to occur.

Many different histone H3 cleavage sites have been reported in the past, even within one study. In differentiating mESC [184], A21 is the primary site of cleavage, though multiple other sites were also found at T22, K23, A24, R26 and K27. By the use of Western blotting, Edman degradation and MS respectively, we confirmed two of those cleavage sites in hESC, namely A21 and R26, and assigned residue 31 as an additional new cleavage site in hESC. Unfortunately, with the techniques applied, no definite distinction could be made between the H3.1/H3.2 or H3.3 isoform cleavage respectively. Nevertheless, we expect H3.1 or H3.2 to be the cleaved isoform, since A31 is followed by threonine just as A21, which can thus be suspected to be susceptible to a similar enzyme activity.

It is worth noting that despite the many hurdles that need to be overcome when using MS for the analysis of histone H3, several arguments add up to the likelihood of this newly identified cleavage site at residue 31. First, the annotated peptide was N-terminally propionylated, indicating no in-source decay causing this fragment to appear. Second, elution time patterns of the precursor and its cleaved form are distinct, further arguing against in-source formation of the cleaved fragment out of the intact precursor. Third, by first separating the histones with SDS-PAGE, the location of annotated histone H3 and its cleaved form in the gel allows to project the peptide data back onto the precursor proteins: the N-terminally cleaved fragment could only be annotated in lower MW (cH3) gel bands, in contrast to the N-terminal peptide itself which was found only in the highest MW (intact H3) fraction. Finally, all spectra corresponding to a cleaved peptide were manually validated by an expert before taken into account. On the other hand, more cleavage sites are expected to be present which cannot be identified by MS, as seen for the A21 and R26 sites detected by a specific antibody and Edman degradation respectively. This can be explained by the limited possibilities of MS to annotate small peptides with high reliability. When a cleavage site is found to be close to an arginine, which is cleaved during digestion, the resulting peptides will consist of too little Aa to be identified reliably. As a result not all possible cleavage sites can theoretically be identified by MS but their existence should not be disregarded.

Only few of the published histone clipping reports actually also categorize the protease responsible for this event (reviewed in [204]). Duncan and colleagues (2008) assigned the cleavage to cathepsin L, a lysosomal cysteine protease, whilst glutamate dehydrogenase has been brought forward by Mandal et al. (2013). Santos-Rosa et al. (2009) who originally published the yeast clipping event, could only categorize the enzyme as a serine protease. It was just recently that Xue and colleagues [205] found that the vacuolar protein Prb1 (Cerevisin) is required for the N-terminal H3 clipping in *Saccharomyces*

cerevisiae. Finally, the Foot-and-Mouth disease virus (FMDV) expresses the so-called protease 3C, a cysteine protease, in the host cells which mediates clipping of host histone H3 at leucine 20 [187,188].

The incubation assays performed here with the application of different inhibitors point out that the clipping enzyme is a serine protease, since apart from AEBSF none of the inhibitors for other protease classes could establish any effect on the incubation. This was confirmed by the fact that the PIC, which inhibits mainly serine and cysteine proteases, was also able to inactivate the clipping activity. We also included three specific inhibitors for cathepsin L, in analogy to the clipping in mESC, but no decrease in cleavage was seen due to these inhibitors. However, caution should be taken in the search for the responsible enzyme. As also others already suggested [184,191], enzyme redundancy and overlapping functions could impede its identification.

Despite the great epigenetic promise of such radical PTM, histone clipping and more specifically its biological potential and the mechanisms by which it could exert its transcriptional effects, remain surprisingly understudied. Since the two landmark discoveries of H3 clipping in mouse and yeast in 2008 and 2009 respectively, four major mechanisms have been formulated by which histone proteolysis can influence gene expression programs [185].

First, in yeast, a direct regulatory role of gene expression has been attributed to the removal of the N-tail and its repressive marks at promoter regions [191]. As such, histone H3 cleavage clears repressive marks massively, hence allowing for gene expression activation. Second, cH3 might provide a new binding site for protein complexes that could not be bound before clipping, thus fulfilling an active role in protein recruitment. On the other hand, also a passive regulatory role should be considered if other proteins are no longer able to bind the shortened histone H3 [185,191]. Third, Santos-Rosa and colleagues [191] also proposed nucleosome eviction and histone replacement as another framework in which clipping might regulate gene expression, as they report that H3 clipping precedes nucleosome eviction and subsequent gene induction. Finally, the N-terminal peptide itself might establish its translation regulation by binding its own mRNA [185]. Which of these mechanisms, if any, is at play in (differentiating) hESC remains to be elucidated.

In short, based on previous findings and theoretical background, H3 clipping could be involved in several processes linked to gene expression control and differentiation. But the question still remains as to whether H3 proteolysis correlates directly with or causes such processes. Further studies will hopefully help elucidate the role of this new epigenetic mark in hESC and their differentiation process.

4.6. Conclusions

As a member of the epigenetic network, regulated histone proteolysis has been occasionally described earlier in diverse biological settings, yet being largely understudied. Histone H3 clipping and histone proteolysis in general might skew experimental findings substantially, both from a biological and a technical point of view. In this report we show for the first time that this PTM is also present in human ESC, and is mediated by a serine protease. The temporal pattern of cleaved H3 is highly dependent on the culture protocol applied, as seen for both cell lines used in this report. Although in first instance we also detected the clipping upon early stem cell differentiation, we found that the clipping process is not necessarily accompanied by a decrease in Oct4 expression, as also undifferentiated hESC can contain cleaved H3 fragments. Thus, more research is needed to fully elucidate the potential biological role(s) of histone H3 cleavage.

"An expert is a person who has made all the mistakes that can be made in a very narrow field."

— Niels Bohr

CHAPTER 5:

PITFALLS IN HISTONE PROPIONYLATION DURING BOTTOM-UP MASS SPECTROMETRY ANALYSIS

Adapted from "**Pitfalls in histone propionylation during bottom-up mass spectrometry analysis**"

Paulien Meert, Elisabeth Govaert, Ellen Scheerlinck, Maarten Dhaenens* and Dieter Deforce*

Proteomics 2015, 15, 2966–71

* Equal contribution

5. PITFALLS IN HISTONE PROPIONYLATION DURING BOTTOM-UP MASS SPECTROMETRY ANALYSIS

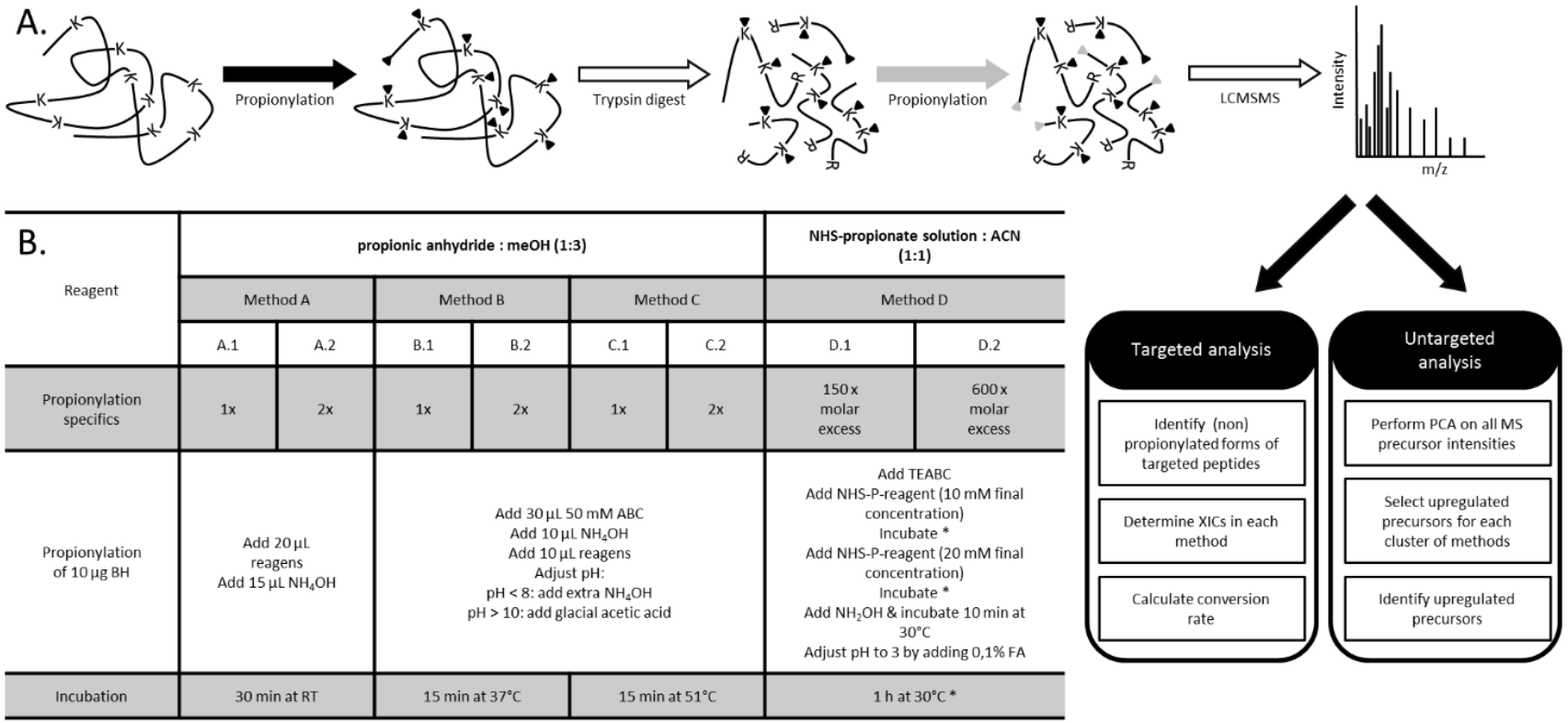
5.1. Abstract

Despite their important role in regulating gene expression, hPTMs modifications remain technically challenging to analyze. For identification by bottom-up MS, propionylation is required prior to and following trypsin digestion. Hereby, more hydrophobic peptides are generated enabling RP-HPLC separation. When histone dynamics are studied in a quantitative manner, specificity and efficiency of this chemical derivatization are crucial. Therefore we examined eight different protocols, including two different propionylation reagents. This revealed amidation (up to 70%) and methylation (up to 9%) of carboxyl groups as a side reaction. Moreover, incomplete (up to 85%) as well as aspecific propionylation (up to 63%) can occur, depending on the protocol. These results highlight the possible pitfalls and implications for data analysis when doing bottom-up MS on histones.

5.2. Background

Histones are subjected to a diverse array of PTMs, thereby regulating the accessibility of the underlying DNA, effecting both physiology and disease [63,206]. MS has become a powerful tool to simultaneously identify and quantify these PTMs. Nevertheless, sample preparation for bottom-up MS strategies is complicated by the requirement of chemical derivatization such as propionylation prior to and following trypsin digestion (Figure 5.1.A) [128]. This step modifies all free primary amine groups (the N-termini and the ε-amino group of unmodified and monomethylated lysine (K)), hence changing the tryptic into Arg-C specificity and resulting in larger (6-20 Aa), more hydrophobic and readily identifiable peptides. However, in order to optimize identification and study histone dynamics in a quantitative way, this propionylation reaction has to be specific as well as efficient. Only then, accuracy and reproducibility can be guaranteed. Propionylation specificity of histones can be hampered by the high abundance of hydroxyl containing residues (serine (S), threonine (T) and tyrosine (Y)) which can be aspecifically propionylated in addition to the primary amines, hereafter referred to as "overpropionylation". On the other hand, an efficient reaction implies that all free primary amine groups should react, leaving no peptides "underpropionylated". Finally, unanticipated side-reactions can hinder the peptide annotation rate and bias quantification.

Based on the protocols most often used in current literature we developed 4 different propionylation methods (A to D) comprising two different types of propionylation agents, several buffer types and incubation temperatures (Figure 5.1.B) [128,179,198,207,208]. The methods using propionic anhydride were performed with a single as well as a double round of propionylation pre- and post-digestion, to monitor the advantage of an extra round of propionylation. For similar reasons method D was carried out using two different concentrations of NHS-propionate. Each protocol was performed in triplicate on 10 µg bovine histones and subsequently 1 µg of each sample was analyzed by MS, using a label-free information-dependent acquisition strategy on a TripleTOF 5600 (Sciex). In a first targeted data analysis step on this dataset, eight different manually validated peptides were monitored throughout all eight methods to gain a first insight into the conversion rate, specificity and efficiency of each propionylation protocol. In the subsequent untargeted evaluation strategy, a PCA analysis (Progenesis QI, NonLinear Dynamics, Waters) on all MS precursor intensities present, 11.247 in total, was performed in order to verify clustering (and thus reproducibility) of the experimental conditions and to check for outliers without prior knowledge of these peptides' identity. For each cluster of methods MS precursors were selected that were significantly (p-value ≤ 0.0001) most abundant (highest mean) within this cluster. Each group of extracted MS precursors was then subjected to consecutive rounds of searches to define the occurrence of any unanticipated side-reactions (Figure 5.1.A).



B.

Reagent	propionic anhydride : meOH (1:3)						NHS-propionate solution : ACN (1:1)	
	Method A		Method B		Method C		Method D	
	A.1	A.2	B.1	B.2	C.1	C.2	D.1	D.2
Propionylation specifics	1x	2x	1x	2x	1x	2x	150 x molar excess	600 x molar excess
Propionylation of 10 µg BH	Add 20 µL reagens Add 15 µL NH_4OH		Add 30 µL 50 mM ABC Add 10 µL NH_4OH Add 10 µL reagens Adjust pH: pH < 8: add extra NH_4OH pH > 10: add glacial acetic acid				Add TEABC Add NHS-P-reagent (10 mM final concentration) Incubate * Add NHS-P-reagent (20 mM final concentration) Incubate * Add NH_2OH & incubate 10 min at 30°C Adjust pH to 3 by adding 0,1% FA	
Incubation	30 min at RT		15 min at 37°C		15 min at 51°C		1 h at 30°C *	

Figure 5.1 Propionylation workflow and overview of the different protocols

A: Propionylation is carried out prior to (▼) as well as post digestion (▼), followed by LCMSMS analysis. During the first propionylation reaction K, monomethylated K and the protein N-terminus become derivatized. After digestion the newly generated peptide N-termini get propionylated as well. The generated LCMSMS data was analyzed using two approaches: **(i)** left box: a targeted approach, defining the conversion rate based on identified peptides **(ii)** right box: an untargeted approach, based on differential MS precursor intensities in between methods. The first strategy can be used to determine efficiency and specificity of the protocol, the latter is used to monitor for unexpected side reactions that differ in between protocols. **B**: Table representing the differences between the propionylation methods. The 4 different methods vary in propionylation reagent, buffer and reaction temperature. Methods A to C are performed with a single (1x) as well as a double (2x) round of propionylation before and after digestion, marked as method X.1 and X.2 respectively. Method D is carried out with a 150 times molar excess (method D.1) and a 600 x molar excess (method D.2) of NHS-propionate to the bovine histones.

5.3. Methods

An additional overview of all propionylation methods used and mentioned throughout this work is given in Chapter 11 (Addendum).

Propionylation method A

Twenty µL propionylation reagent (propionic anhydride:methanol 1:3 (v/v)) was added to 10 µg vacuum-dried bovine histones (Roche), immediately followed by the addition of 15 µL ammonium hydroxide. The reaction was incubated at room temperature for 30 min and evaporated in a SpeedVac to remove any volatile remnants. In method A.2 a second round of propionylation was performed prior to trypsin digestion, in contrast to method A.1 where propionylated proteins were immediately reconstituted for digestion. To propionylate the newly generated N-termini, propionylation was also carried out after enzymatic digestion, with the same protocol as before digestion.

Propionylation method B and C

Bovine histones (10 µg) were resuspended in 15 µL H_2O and diluted with 15 µL 100 mM ammoniumbicarbonate (ABC). After adding 10 µL ammonium hydroxide to the sample, propionylation reagent (propionic anhydride:methanol 1:3 (v/v)) was prepared and 10 µL immediately added to the histone sample. If needed, an extra amount of ammonium hydroxide or glacial acetic acid was added to keep an ideal pH of 8. Reactions were maintained at 37°C (method B) or 51°C (method C) for 15 min and vacuum-dried, whereupon samples of method B.2 and C.2 were subjected to a second round of propionylation, in contrast to method B.1 and C.1 where propionylated proteins were immediately reconstituted for digestion. To propionylate the newly generated N-termini propionylation was also carried out after enzymatic digestion, with the same protocol as before digestion.

Propionylation method D

Bovine histones (10 µg) were resuspended in 50 mM triethylammoniumbicarbonate (TEAB). A NHS-propionate stock solution (100 mM water: ACN, 1:1 (v/v)) was added to a final concentration of 10 mM whereby the molar excess to the bovine histones was 150x and 600x for methods D.1 and D.2, respectively. Samples were incubated during 1h at 30°C, followed by the addition of the same amount of NHS-propionate stock solution to reach a final concentration of 20 mM and again incubation for 1h at 30°C. To revert propionylated serine, threonine and tyrosine, HA was added to the sample (4 times the molar excess of HA to the NHS ester) and incubated for 10 min at 30°C. The pH was lowered to 3 by adding the required amount of 1% formic acid. Samples were thoroughly dried in a SpeedVac and digested with trypsin, whereupon the newly generated N-termini were also propionylated using the same protocol as before digestion.

Trypsin digestion

Trypsin digestion was identical in all protocols except for the buffer used: in all methods 50 mM ABC was applied, except for method D where 50 mM TEAB was used. In each case, dried out, propionylated histones were reconstituted in the respective buffers, supplemented with 1 mM CaCl2, 5% ACN and digested overnight at 37°C with trypsin (Promega) at a histone/enzyme ratio of 20:1 [198].

LCMSMS

One microgram propionylated peptides were dissolved in 0.1% formic acid in water (buffer A) without any preceding clean up steps, and separated on a PepMap 100 (C18) column (I.D. 75 µm, length 25 cm, particle size 5 µm) by means of trap-elute injection. An U3000 LC-system (Dionex) was used at a flow rate of 300 nL/min. Elution was performed with 80% acetonitrile/0.1% formic acid (buffer B) using a gradient of 4% to 55% buffer B in 60 min. Quality control samples (QC) contain equal amounts of each sample and 1 µg in total. A TripleTOF 5600 (Sciex) was used for information dependent analysis, using a nano-ESI source in positive ion mode. Survey MS scans were acquired (m/z 400-1250) and up to 20 precursors (m/z 65-2000) with charge state 2+, 3+ or 4+ exceeding the signal threshold (750 counts per second) were isolated for fragmentation by collision induced dissociation, using the dynamic collision energy profile as suggested by the manufacturer.

Data analysis

Targeted as well as untargeted data analysis was performed on triplicate runs of each protocol. In this data set 87% of the MS precursors have a power > 0.8, justifying the sample size of 3 replicates to find a significant difference between groups.

Database searching was performed against a bovine database obtained from the National Center for Biotechnology (NCBI) database, merged with the contaminant database included in the Sciex software containing prevalent contaminations such as keratin and trypsin. During the targeted approach, generated wiff-files (each method performed in triplicate) were thoroughly searched using Protein Pilot 4.5 (Sciex), implementing the following special factors: purified histones, propionylation pre-digestion and propionylation post-digestion. Enzyme specificity was set to Arg-C and an FDR analysis was performed. The data were exported to Excel and filtered using 5% distinct peptide local FDR as cut-off value. The different peptide forms of eight peptides (underpropionylated, desired, overpropionylated) were isolated and XICs of these forms throughout all runs were made using PiekView 1.2. To confirm correct annotation and localization of the propionyl groups, the same dataset was also searched using a Mascot 2.5 in house server (Matrix Science) and only spectra assigned to the same peptide by both ProteinPilot and Mascot were taken into account. Mascot parameters were the following: mass error

tolerances for the precursor ions and its fragment ions were set at 15 ppm and 0.01 Da respectively, enzyme semi-specificity was set to Arg-C, allowing for up to two missed cleavage sites. Variable modifications included acetylation, dimethylation, trimethylation and propionylation on lysine (K), methylation and dimethylation on arginine (R) and oxidation of methionine. Lysine monomethylation was searched as the sum of propionylation and methylation since monomethylated lysine residues can still be propionylated. Propionylation was set as a variable modification at the N-terminus, serine (S), threonine (T) and tyrosine (Y). Only peptides with an expectancy value < 0.01 were considered correctly annotated. Moreover, all spectra used for peptide identification were manually validated. Most peptides were covered throughout the whole Y-ion series, enclosing the Aa susceptible for propionylation, which implies correct localization of the propionyl group. We support this with a representative example in Figure 5.6. MSMS spectra of peptide YQKSTELLIR are depicted. For this peptide three possible propionylation sites (K, S and T) are located next to each other, thereby increasing the risk of incorrect localization.

The propionylation conversion rate was calculated as follows: XIC desired products / (XIC desired products + XIC underpropionylated products + XIC overpropionylated products).

The untargeted approach was carried out using Progenesis QI software (Nonlinear Dynamics, Waters). After alignment of the runs, the MS precursors were filtered, based on retention time (Rt 15-75 min) and charge state (2+-5+) and the data were normalized to all precursors. A multivariate statistical analysis was performed on all 11.247 MS precursors, without any prior peptide identification. The Principal Component Analysis (PCA) in Progenesis uses MS precursor abundance levels across runs to determine the principle axes of abundance variation. Transforming and plotting the abundance data in principle component space separates the samples according to abundance variation. PCA analysis was carried out to check for outliers and see if experimental conditions group together. The MSMS spectra of the differential precursors, significantly most abundant in a cluster of methods (ANOVA p-value < 0.0001; q-value < 2e-6) were exported as an MGF peaklist and searched using a Mascot 2.5 in-house server (Matrix Science). An error tolerant search against a bovine histone database containing prevalent contaminants was performed, with propionylation of lysine and the N-terminus as variable modifications. Mass error tolerances for the precursor ions and its fragment ions were set at 15 ppm and 0.05 Da respectively. Enzyme specificity was set to Arg-C, allowing for up to two missed cleavage sites and results were filtered for confident hits using a 0.01 expectancy cutoff. The most abundant modification was identified for each separate *.mgf file using the "modification statistics" functionality of the 2.5 Mascot server. Next, the relative abundance of this specific modification was calculated for

each cluster separately by dividing the total number of error tolerant modification matches by the number of peptide matches above identity threshold.

5.4. Results

First we compared the average propionylation conversion rate of 8 different peptides in the triplicate runs of each protocol. The chosen peptides are both non-modified as well as biologically modified and originate from histone H3 and H4. Since acetylation on peptide KQLATKAAR results in two isobaric co-eluting forms (H3K18Ac and H3K23Ac) that cannot be distinguished on MS level, we here refer to these isobaric species as KQLATKAAR+Ac. To determine this conversion rate for a specific peptide, the peak area of the extracted ion chromatogram (XIC) for the desired product was divided by the sum of the peak areas representing the total pool of this peptide: the desired form, overreacted products (overpropionylated) and incomplete products (underpropionylated) Figure 5.2.A and Figure 5.2.B). Method A.2 performs best for all eight peptides with an average conversion rate between 93 and 100%. Methods B and C on the other hand have average conversion rates lower than 70% for 7 out of 8 peptides (Figure 5.2.C). The one outlier with a conversion rate of 99% in these protocols coordinately introduces the notion of sequence-dependent propionylation efficiency. This phenomenon can also be seen in the methods using NHS chemistry, where the conversion rate is over 80%, except for the peptides DAVTYTEHAKR and K(Me)SAPATGGVKKPHR where it stays below 65%.

The low conversion rates of methods B to D can either be due to aspecific overpropionylation, inefficient propionylation or a combination of both. Hence, we calculated the average contribution of overpropionylation as well as underpropionyation for all peptides in each protocol based on XICs. When using NHS-chemistry, overpropionylation is the main reason for a low conversion rate, with a peak of up to 60% overpropionylation of peptides DAVTYTEHAKR and K(Me)SAPATGGVKKPHR (Figure 5.3.A). Methods B and C on the other hand mainly suffer from underpropionylation due to incomplete reaction, thereby hampering conversion (Figure 5.3.B). Method A.2 is both specific and efficient for all peptides, which explains the high conversion rate mentioned before.

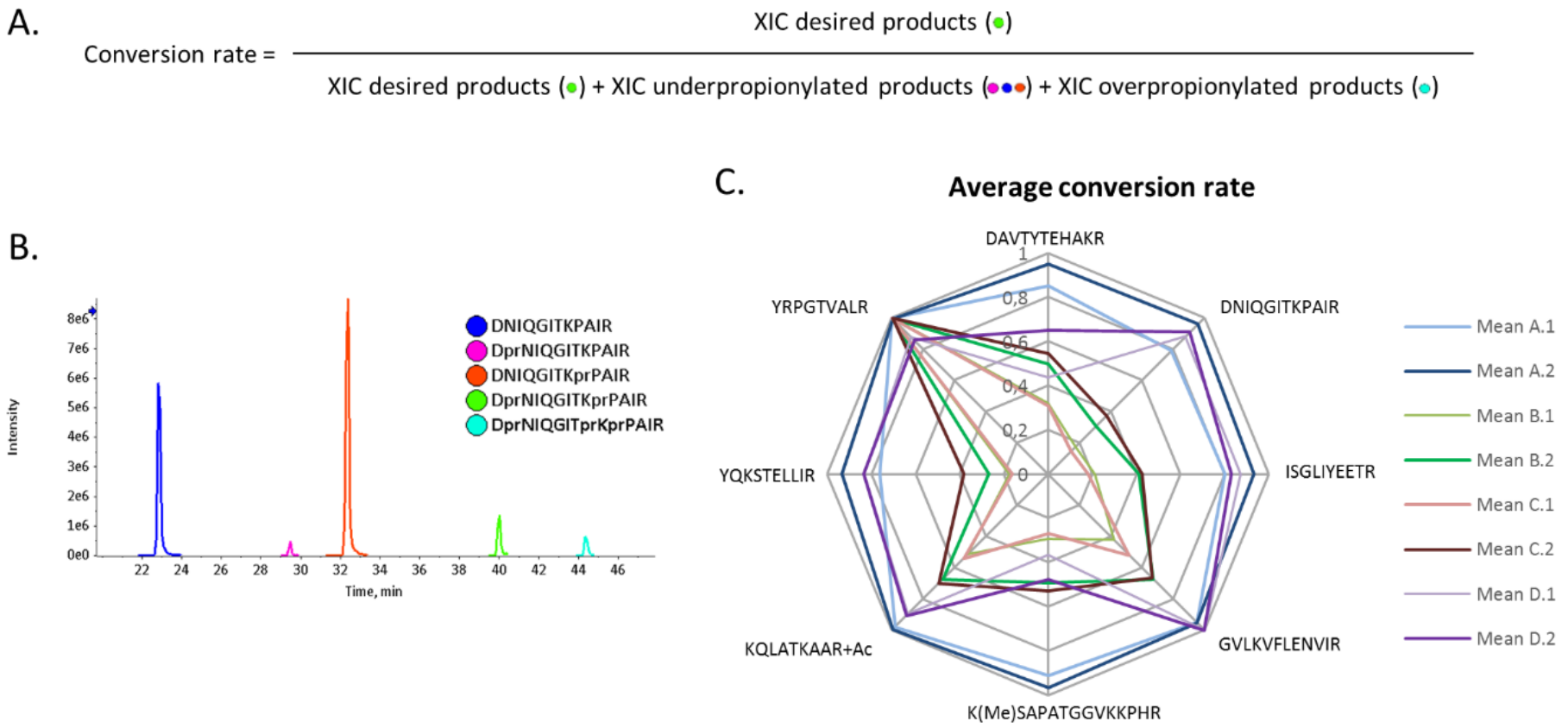


Figure 5.2 Targeted data analysis

A: Formula to determine the conversion rate of a peptide, based on XICs of identified forms. **B**: Composite representation of the XICs of five different forms of peptide DNIQGITKPAIR, generated after the propionylation (pr) workflow: underpropionylated products (● ● ●), desired products (●), overpropionylated products (●). This clearly illustrates the increasing retention that is induced by propionylation. **C**: Radar chart representing the average conversion rate for eight targeted peptides. Each peptide is located on one angle of the radar chart and each method is represented by another color. The conversion rate for each peptide using the different methods is shown on the radius, whereby a conversion rate of 0 is located in the center, increasing outwards.

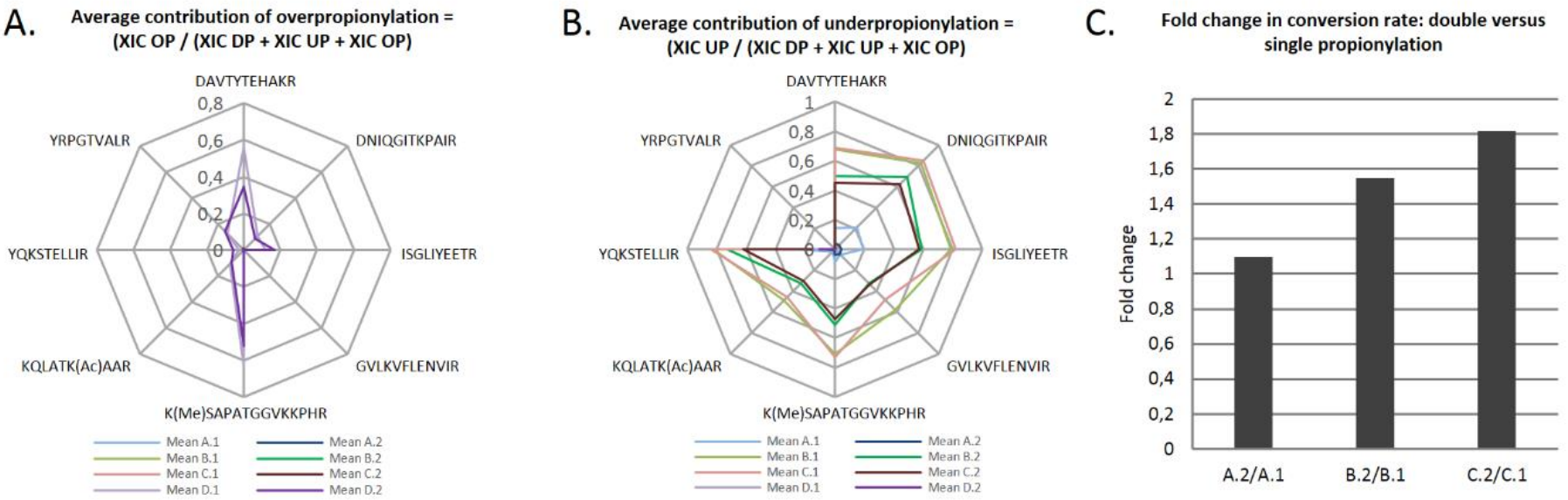


Figure 5.3

Radar chart showing the average contribution of over- (**A**) and underpropionylation (**B**). DP = desired product; UP = underpropionylated product; OP = overpropionylated product for the eight targeted peptides in each protocol. (**C**) Graph showing the fold change in conversion rate when a double round of propionylation is carried out prior to and post digestion, as opposed to a single propionylation.

Next we focused on the effect of a second round of propionylation prior to as well as post digestion on the conversion rate. Therefore we calculated the average increase in conversion rate for all six peptides when using method X.2 instead of method X.1 (Figure 5.3.C). These results indicate that a second round of propionylation increases the conversion rate by lowering underpropionylation without any increase in overpropionylation. This effect is most notable for method B and C with an increase of the conversion rate by a factor of 1.55 for the protocol at 37°C and up to a factor 1.82 for the protocol at 51°C. When using method A an increase by only a factor of 1.10 was observed. This can be explained by the high conversion rate that was already found when performing one propionylation round. Of interest, technical variation of propionylation in between triplicates as well as sequence-dependent propionylation in between the peptides subjected to the same protocol, lowered considerably when performing a second round of propionylation (Table 5.1). Especially method A.2 thus shows little sequence-dependent propionylation at all, with average conversion rate in between 93 and 99% for all six peptides as opposed to 79 to 99% when only one round of propionylation was applied (method A.1). Thus, a second round of propionylation increases the conversion rate and thereby lowers technical variation as well as sequence-dependent propionylation whereby method A.2 emerges as the best candidate protocol for subsequent quantitative MS analysis.

Table 5.1
Each propionylation method was performed in triplicate. The average propionylation conversion rate is shown for each targeted peptide, together with the standard deviation.

Peptide conversion rate																
Peptide	Method A.1		Method A.2		Method B.1		Method B.2		Method C.1		Method C.2		Method D.1		Method D.2	
	Average	SD	Average	SD	Average	SD	Average	SD	Average	SD	Average	SD	Average	SD	Average	SD
DAVTYTEHAKR	0,85	0,08	0,95	0,04	0,32	0,12	0,50	0,08	0,31	0,03	0,54	0,01	0,44	0,04	0,65	0,07
ISGLIYEETR	0,80	0,13	0,93	0,02	0,21	0,07	0,41	0,13	0,18	0,08	0,43	0,08	0,87	0,02	0,83	0,02
GVLKVFLENVIR	0,95	0,03	0,95	0,05	0,42	0,27	0,67	0,08	0,52	0,16	0,67	0,08	1,00	0,00	1,00	0,00
DNIQGITKPAIR	0,79	0,19	0,96	0,03	0,17	0,08	0,31	0,13	0,15	0,06	0,37	0,06	0,88	0,02	0,91	0,02
YQKSTELLIR	0,76	0,11	0,94	0,02	0,17	0,06	0,27	0,04	0,16	0,08	0,38	0,03	0,83	0,08	0,84	0,04
YRPGTVALR	1,00	0,00	0,99	0,00	0,99	0,00	0,99	0,00	1,00	0,00	1,00	0,00	0,88	0,05	0,86	0,02
KQLATKAAR+Ac	0,98	0,01	1,00	0,01	0,51	0,12	0,68	0,08	0,54	0,06	0,70	0,06	0,90	0,04	0,91	0,03
K(Me)SAPATGGVKKPHR	0,91	0,03	0,97	0,02	0,29	0,10	0,49	0,11	0,27	0,06	0,53	0,08	0,37	0,06	0,48	0,04

While of obvious value in finding the best possible sample preparation protocol, these analyses are all based on preceding peptide identification, and are therefore targeted. Nevertheless, it is possible that unanticipated side reactions occur, thereby generating peptides which remain unidentified using standard search parameters. In order to search for these unanticipated side reactions a "quantify-then-identify" strategy was applied under the form of an "MS precursor intensity based" PCA (Progenesis QI, Nonlinear Dynamics, Waters) on the total of 11.247 MS precursors, (Figure 5.4).

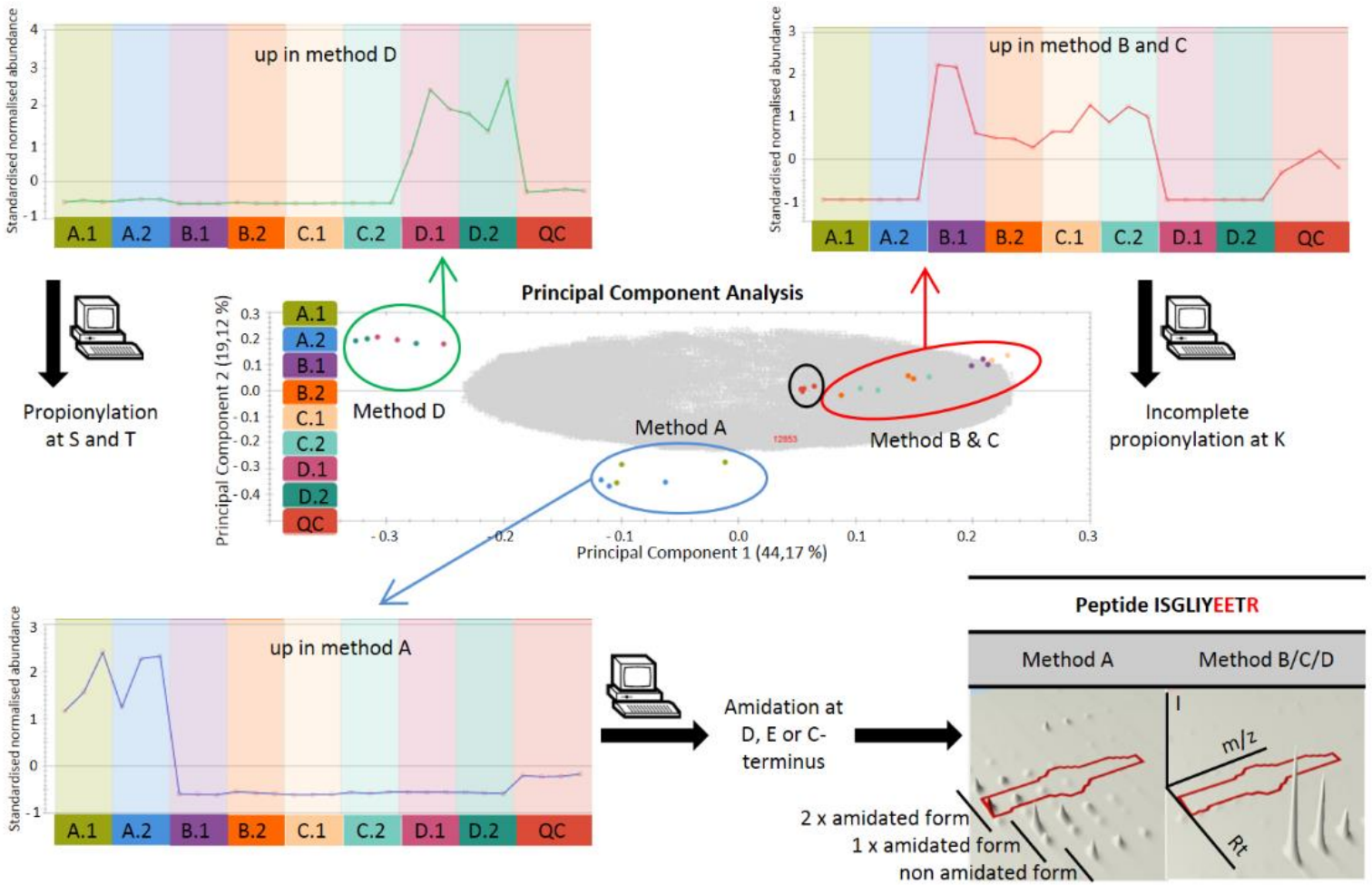


Figure 5.4 Untargeted, precursor-based data analysis

PCA was performed on MS precursor intensities from a label-free IDA analysis of triplicate experiments and four QC samples resulting in the clustering of method A, method D and methods B & C. Precursors with an ANOVA p-value ≤ 0.0001 and significantly most abundant for one cluster of methods were filtered. A representative abundance profile of a differential precursor is shown for each cluster (method A: blue; method B and C: red; method D: green). Identification of the MSMS spectra linked to these exported precursors confirmed that there is an enrichment of overpropionylated peptides in method D, and that underpropionylation is mainly found in methods B and C. A new modification was revealed to be enriched in method A: amidation of D, E and the C-terminus. This can result in a dispersion of precursor signal intensity over the generated peptide forms, illustrated by peptide ISGLIYEETR. The amino acids susceptible for amidation are highlighted in red and the different peptide forms are marked.

Four quality control (QC) samples were included, next to the triplicate method samples, resulting in a total of 144.784 MSMS spectra generated over all different runs. QC samples are identical and contain equal amounts of each sample and 1 µg in total. Because they contain all possible precursors within one sample, they can be used as a precursor alignment template for the Progenesis QI software, which is very important here because considerable differences were induced by the different protocols. Since this PCA is carried out using MS precursors instead of identified peptides, no prior knowledge concerning annotation parameters is required. The aggregation of the different methods in this PCA analysis (87% of the MS precursors has a power > 0.8) shows three different clusters: method A, method B & C and method D with PC1 explaining 44% and PC2 19% of the variation. MSMS spectra from precursors that are significantly most abundant for a cluster of methods were extracted into a separate *.mgf file (ANOVA p-value < 0.0001; q-value < 2e-6), generating three clusters of MS precursors: most abundant in method A, most abundant in method B & C and most abundant in method D. These three separate *.mgf files (comprising 21.406, 37.736 and 36.141 MSMS spectra respectively) were then each subjected to an error tolerant search with both N-terminal and K-specific propionylation set as variable modification (Mascot 2.5, Matrix Science). The most occurring error tolerant modification was identified for each cluster, using the modification statistics (Figure 5.5).

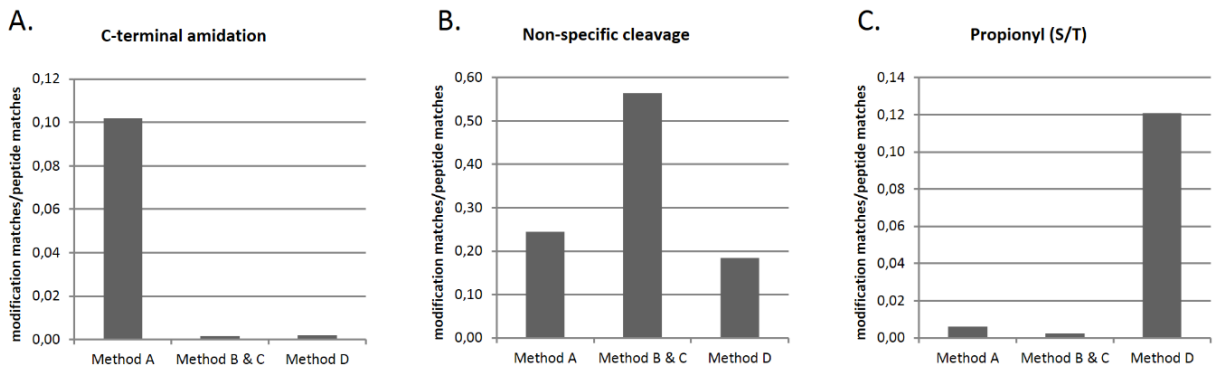


Figure 5.5 Graphs representing the relative abundance of a modification
After PCA, the most occurring error tolerant modification was identified within all three "method clusters". Next, we also monitored the presence of each of these modifications in the other clusters. Shown here is the relative percentage of annotated spectra that carry: (**A**) amidation of D, E and the C-terminus, (**B**) non-specific cleavage and (**C**) aspecific propionylation of S and T.

Indeed, *in vitro* induced modification significantly outnumbered biologically relevant PTMs such as acetylation. Earlier findings of the targeted analysis were confirmed herein: **(i)** precursors that were most abundant in method D were mainly identified as overpropionylated (propionylation on S and T as the most identified error tolerant modification), **(ii)** precursors that were most abundant in method B and C mainly suffer from underpropionylation. The latter group thus also comprised "semi-ArgC" peptides, in

which the protein was "non-specifically" cleaved at a K because this amino acid was not sufficiently propionylated in the reaction prior to the trypsin digest. Surprisingly, the third group of differential precursors **(iii)** which were most abundant in method A, also shared a common PTM: amidation (-0.9840 Da) on aspartic (D) and glutamic acid (E), as well as on the C-terminus. These identifications (MSMS spectra of both the amidated and non-amidated peptide form) are shown in Figure 5.6.

Remarkably, the reactivity of carboxyl groups in method A is not only limited to amidation. Also methylation of D, E and C-terminus can occur when mixing propionic anhydride with methanol. Yet, this side reaction was not identified for method B and C even though methanol was used as well. In order to estimate the impact of these side reactions (amidation and methylation of COOH-groups) on quantification we calculated the relative abundance of both the amidated and carboxy methylated peptide form compared to its non-reacted counterpart (Figure 5.7).

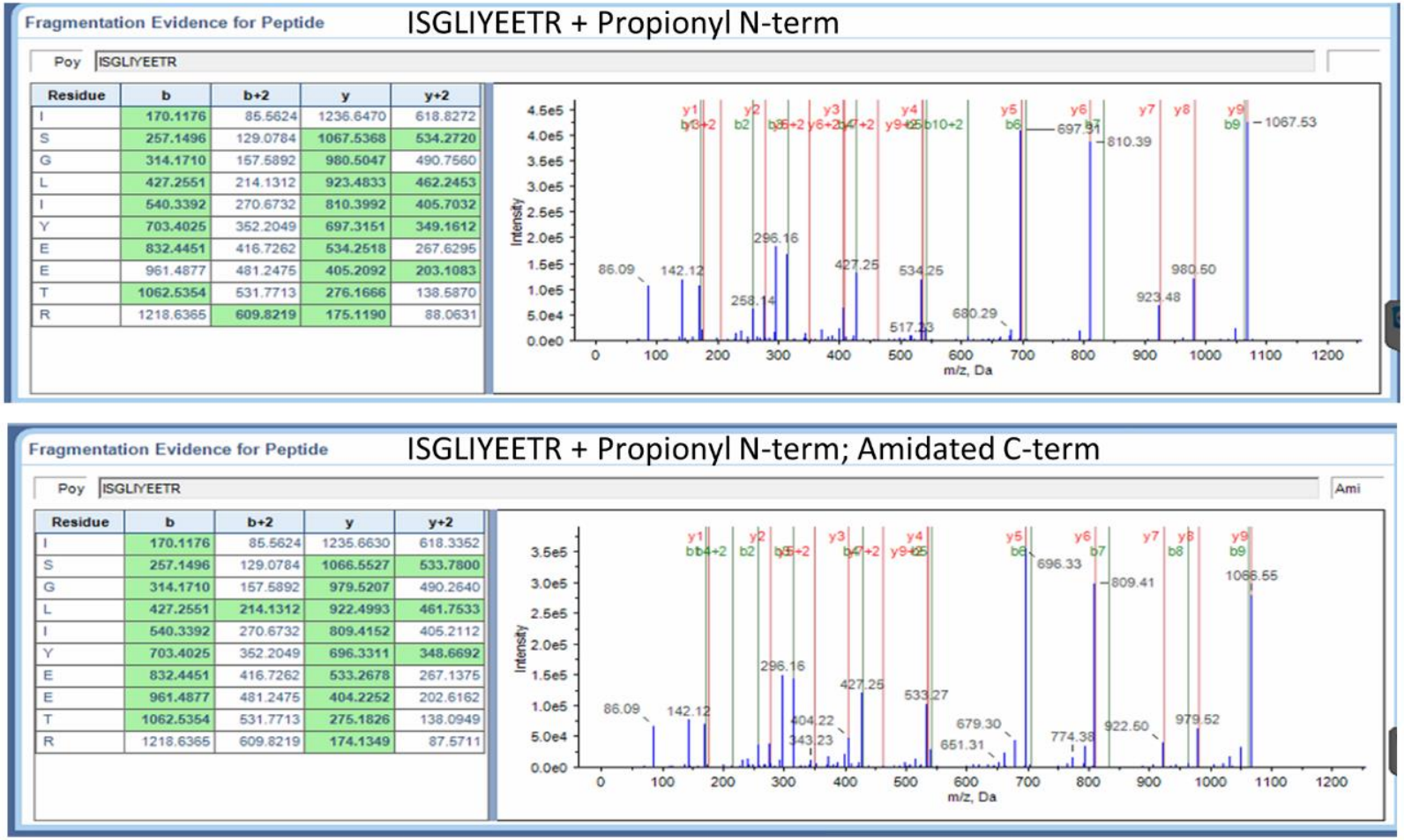


Figure 5.6 C-terminal amidation is a well-supported identification

The upper panel shows the MSMS spectrum of propionylated peptide ISGLYEETR, its C-terminal amidated counterpart is depicted in the lower panel, displaying a complete y-ion series 0.98 Da below the original series.

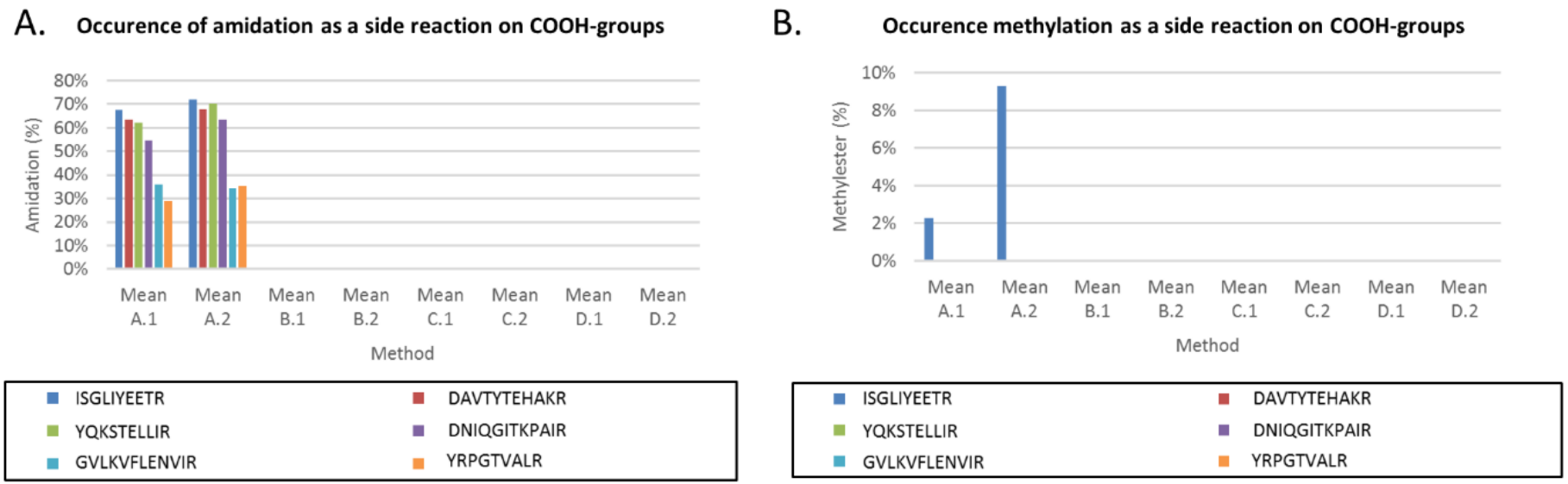


Figure 5.7 Occurrence of unwanted amidation and methylation of carboxyl groups

Amidation (**A**) and methylation (**B**) of a carboxyl groups can occur as a side reaction when using method A. Amidation is present on all peptides depicted, in contrast to the formation of a methylester which was identified for peptide ISGLIYEETR only.

All 8 peptides used for targeted analysis were investigated, yet 2 peptides were not included in Figure 5.7. Especially for peptide KQLATKAAR+Ac, accurate quantitation was impaired by the fact that the amidated form elutes first and its naturally occurring first isotope co-elutes with the non amidated peptide precursor mass, hindering XIC based quantification (Figure 5.8). For the other peptide, K(Me)SAPATGGVKKPHR, amidation nor methylation was identified. As shown in Figure 5.7, amidation at COOH occurs far more frequently and intense than methylation in this protocol. The latter was only detected for peptide ISGLIYEETR and no more than 9% was affected. Amidation on the other hand can convert up to 70% of a peptide in an amidated form, depending on the presence of D and E in the sequence.

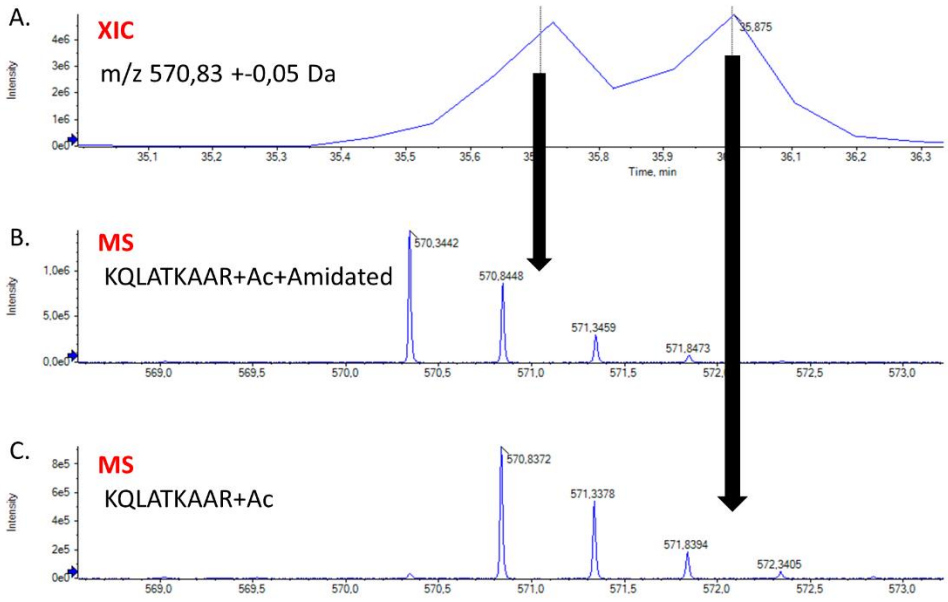


Figure 5.8
Co-elution of peptide KQLATKAAR+Ac (m/z 570.83) (**B**) with its amidated counterpart (**C**) hampers correct quantitation by means of XIC (**A**).

As shown here, each method has its own shortcomings. However, the most often used metric for studying biology is the relative abundance of PTMs in which the intensity of modified peptides are expressed relative to all peptides sharing that same sequence. Thus, we quantified the relative abundance of one such PTM, H3K23ac, within the bovine histone sample in all different protocols. Hereby we assumed the researcher to be blind to an unexpected side reaction, underpropionylation or overpropionylation. This was estimated as the percentage of peptide KQLATKAAR+Ac by all peptides sharing that same sequence (side reactions, underpropionylation or overpropionylation not taken into account). As shown in Table 5.2, estimations only vary slightly (between 28% and 35%). While this is reassuring for the conclusions on biology reported to date, it emphasizes the importance of calculating

the relative abundance in each run before comparing the samples, as opposed to directly comparing precursor intensities in between separate runs. Also, the standard deviation of the relative abundance in protocols using only a single round of propionylation cautions for the limited accuracy of estimation for these methods. Equally important however, abundant side reactions such as amidation not only reduce the signal of the *in vivo* relevant precursor ion, they also generate large amounts of uninformative new precursors that are being selected for MSMS during the DDA acquisition, as these new forms have a different retention time (Figure 5.4). It is therefore advisable to check for side reactions (amidation / methylation at COOH), underpropionylation or overpropionylation. These peptides can then be taken into account when reporting on the total amount of features detected and identified.

Table 5.2 The relative abundance (%) of acetylation at H318-26 is shown as estimated by each protocol (average + SD)

Estimated relative abundance of acetylation at H3 18-26		
Method	**Average abundance (%)**	**SD (%)**
Method A.1	34%	8,34%
Method A.2	28%	1,21%
Method B.1	31%	4,44%
Method B.2	31%	0,39%
Method C.1	35%	0,53%
Method C.2	31%	0,70%
Method D.1	31%	0,91%
Method D.2	33%	1,69%

5.5. Conclusion

In conclusion, this dataset pointed out that a second round of propionylation increases the conversion rate as well as the reproducibility of precursor quantification and is therefore strongly recommended. Nevertheless, several pitfalls in propionylating histones for bottom-up MS were disclosed: incomplete derivatization, aspecific propionylation and side reactions on carboxyl groups. Each of these events has its own implications during data analysis (Table 5.3). When focusing on identification of histones the following should be taken into account: using method A, amidation and methylation of COOH-groups should be added as a variable modification; method B and C will benefit from allowing non-specific cleavage at K (or using trypsin as enzyme in the search parameters with a high number of missed cleavages) and setting N-terminal propionylation as variable modification instead of fixed; adding propionylation on S, T and Y will increase identifications when using method D. For accurate quantification, we strongly recommend to use relative abundances, as these appear reproducible between different protocols and can thus be considered the most robust option. Amidation was only found as a side reaction thanks to an untargeted evaluation strategy based on a PCA on MS precursor intensities and identification of differential MS precursor abundance in between protocols. Therefore,

we would like to stress the importance of using such an approach and recommend including it when comparing or evaluating other protocols.

Table 5.3 Table summarizing the pitfalls for each method and suggestions for data analysis

Method	Pitfall	Suggestions for data analysis				
		Peptide identification				Quantification of histone modification
		Amino Acid	Modification	Variable	Fixed	
A	Amidation at COOH	C-terminus/ D/ E	Amidation	x		Use relative quantification within the same run
	Methylation at COOH	C-terminus/ D/ E	Methylation	x		
	/	N-terminus	Propionylation		x	
B	Underpropionylation	N-terminus/ K/ Kme	Propionylation	x		Use relative quantification within the same run
		K	Non specific cleavage	x		
C	Underpropionylation	N-terminus/ K/ Kme	Propionylation	x		Use relative quantification within the same run
		K	Non specific cleavage	x		
D	Overpropionylation	S/ T/ Y	Propionylation	x		Use relative quantification within the same run
	/	N-terminus	Propionylation		x	

"The most exciting phrase to hear in science, the one that heralds new discoveries, is not 'Eureka!' but 'That's funny …'."

— Isaac Asimov

CHAPTER 6:

EFFICIENT AND SPECIFIC CHEMICAL DERIVATIZATION OF HISTONES BY MEANS OF PROPIONIC ANHYDRIDE

6. EFFICIENT AND SPECIFIC CHEMICAL DERIVATIZATION OF HISTONES BY MEANS OF PROPIONIC ANHYDRIDE

6.1. Introduction

Histone proteins are essential elements for DNA packaging. Moreover, also the PTMs that are extremely abundant on these proteins, contribute in modeling chromatin structure and recruiting enzymes involved in gene regulation, DNA repair and chromosome condensation. Dysregulation of these processes has been intimately associated with the development of diseases such as cancer [209–213]. This fundamental aspect, together with the epigenetic inheritance of histone PTMs, underlines the importance of having biochemical techniques for their characterization. Over the past two decades, significant improvements in mass accuracy and resolution of mass spectrometers, have made LC-coupled MS the strategy of choice for accurate quantification of protein PTMs. Still, the main hurdle to overcome when analyzing histone PTMs is the chemical derivatization of lysine residues prior to this MS analysis. When the technical variation is high and unpredictable, biological conclusion are dubious. In Chapter 5 we investigated the limitations and biases of the most widely adopted sample preparation protocols for histone propionylation. We notice that some protocols lead to incomplete derivatization, while others produce a variety of side products such as amidation, methylesterification and unwanted propionylation on S, T or Y. In this chapter, we further elaborate on these protocols and examine some new approaches, in order to address these previously described challenges during sample preparation.

Several attempts were made to prevent amidation, under- or overpropionylation from happening. Still, most adjustments applied seemed to exchange one problem for another. Finally, reversing overpropionylation by adding hydroxylamine (HA) resulted in an effective protocol with efficient propionylation on primary amines and no unwanted side products remaining. We recommend using this method for future analysis of histones through bottom-up MS.

6.2. Methods

An overview of all propionylation methods used and mentioned throughout this work is given in Chapter 11 (Addendum). For all methods, at least 3 replicates were analyzed.

Propionylation method A.reverse (method A.r)

Ten µg bovine histones were vacuum dried, followed by the addition of 15 µL ammonium hydroxide. Next, 20 µL propionylation reagent (propionic anhydride:methanol 1:3 (v/v)) was added. The reaction was incubated at room temperature for 30 min and evaporated in a SpeedVac to remove any volatile remnants. A second round of propionylation was performed prior to trypsin digestion. To propionylate the newly generated N-termini, propionylation was also carried out after enzymatic digestion, with the same protocol as before digestion.

Propionylation method A.together (method A.t)

Twenty µL propionylation reagent (propionic anhydride:methanol 1:3 (v/v)) was mixed with 15 µL ammonium hydroxide, immediately followed by the addition of this mixture to a pellet of 10 µg vacuum-dried bovine histones. The reaction was incubated at room temperature for 30 min and evaporated in a SpeedVac to remove any volatile remnants. A second round of propionylation was performed prior to trypsin digestion. To propionylate the newly generated N-termini, propionylation was also carried out after enzymatic digestion, with the same protocol as before digestion.

Propionylation method A.no NH_4OH (method A.n)

Twenty µL propionylation reagent (propionic anhydride:methanol 1:3 (v/v)) added to a pellet of 10 µg vacuum-dried bovine histones, immediately followed by the addition of 20 µL water. The reaction was incubated at 37°C for 30 min and evaporated in a SpeedVac to remove any volatile remnants. A second round of propionylation was performed prior to trypsin digestion. To propionylate the newly generated N-termini, propionylation was also carried out after enzymatic digestion, with the same protocol as before digestion.

Propionylation method B.separate (method B.s)

Ten µg vacuum-dried bovine histones were dissolved in 10 µL of 50 mM ABC, immediately followed by 10 µL methanol, 10 µL propionic anhydride. Finally, 20 µL ammonium hydroxide was added to reach a pH above 8. The reaction was incubated at room temperature for 30 min and evaporated in a SpeedVac to remove any volatile remnants. A second round of propionylation was performed prior to trypsin digestion. To propionylate the newly generated N-termini, propionylation was also carried out after enzymatic digestion, with the same protocol as before digestion.

Propionylation method E

Fifty µL propionic acid was added to a pellet of 10 µg vacuum-dried bovine histones, immediately followed by the addition of 5 µL propionic anhydride. The reaction was incubated at room temperature for 6 hours and evaporated in a SpeedVac to remove any volatile remnants. A second round of propionylation was performed prior to trypsin digestion. To propionylate the newly generated N-termini, propionylation was also carried out after enzymatic digestion, with the same protocol as before digestion.

Propionylation method F

Fifteen µL 0.1 M ethanolamine was added to a pellet of 10 µg vacuum-dried bovine histones, immediately followed by the addition of 5 µL propionic anhydride. The reaction was incubated at room temperature for 30 min and evaporated in a SpeedVac to remove any volatile remnants. A second round of propionylation was performed prior to trypsin digestion. To propionylate the newly generated N-termini, propionylation was also carried out after enzymatic digestion, with the same protocol as before digestion.

Propionylation method G

Fifteen µL 0.1 M triethanolamine was added to a pellet of 10 µg vacuum-dried bovine histones, immediately followed by the addition of 5 µL propionic anhydride. The reaction was incubated at room temperature for 30 min and evaporated in a SpeedVac to remove any volatile remnants. A second round of propionylation was performed prior to trypsin digestion. To propionylate the newly generated N-termini, propionylation was also carried out after enzymatic digestion, with the same protocol as before digestion.

Propionylation method H (original – 42x – 20x – 5x)

Twenty µL 1 M TEAB was added to a pellet of 20 µg vacuum-dried bovine histones, immediately followed by the addition of 20 µL propionic anhydride reagent. Of importance, unlike previous methods, isopropylalcohol is used instead of methanol to prevent the formation of methylesters. Depending on which method used, the ratio (v/v) of propionic anhydride and isopropylalcohol differs: 1:3 for method H, 1:79 for method H 42x, 1:170 for method H 20x and 1:680 for method H 5x. Hereby 42x, 20x and 5x represent the molar excess of propionic anhydride compared to the amount of primary amines present in the sample. The reaction was incubated at room temperature for 30 min and afterwards 20 µL water was added for hydrolysis during 30 min at 37°C. Subsequently, the sample was evaporated in a SpeedVac to remove any volatile remnants. A second round of propionylation was performed prior to trypsin digestion for method H, but this step was left out in method H 42x, 20x and 5x. To propionylate the

newly generated N-termini, propionylation was also carried out after enzymatic digestion, with the same protocol as before digestion.

Reversing overpropionylation through boiling

Fifty µL of 50 mM ABC was added to a 5 µg of vacuum-dried bovine histones, which were propionylated using method H 42x. By means of a pH indicator strip the pH was checked and, if necessary, adjusted to 8 using ammonium hydroxide. The reaction was incubated at 99°C for 1 h. After incubation, the solution was evaporated in a SpeedVac to remove any volatile remnants.

Reversing overpropionylation by addition of hydroxylamine

Fifty µL of 0.5 M hydroxylamine was added to 5 µg of vacuum-dried bovine histones, which were propionylated using method H 42x. Around 15 µL of ammonium hydroxide was added for an increase in pH to 12. The reaction was incubated at room temperature for 20 min. After incubation, the pH was lowered to 3 by adding 5% formic acid. Finally, the solution was evaporated in a SpeedVac to remove any volatile remnants.

Trypsin digestion

Trypsin digestion was identical in all protocols, apart from the buffer used: in all methods 50 mM ABC was applied, except for method H and derivatives, where 50 mM TEAB was used. In each case, dried out, propionylated histones were reconstituted in the respective buffers, supplemented with 1 mM $CaCl_2$, 5 % ACN and digested overnight at 37°C with trypsin (Promega) at a histone/enzyme ratio of 20:1 [198].

LCMSMS

Because of technical issues, data acquisition was done by means of two different LCMS setups. Data analysis however was done in a similar way (see Data analysis)

LCMS for samples A.n, B.s, F and G was performed using a nanoACQUITY UPLC system. An analytical column (100 µm x 100 mm nanoACQUITY UPLC 1.7 µm Peptide BEH) was used at a flow rate of 300 nL/min and peptides were separated using the gradient as shown in Figure 6.1. A Synapt G2Si was used for high definition data dependent analysis, using a nano-ESI source in positive ion mode. Survey MS scans were acquired using a fixed scan time of 400 ms. Tandem mass spectra of up to 9 precursors with charge state 2+, 3+, 4+ or 5+ were generated using collision induced dissociation upon exceeding an intensity threshold of 2000 cps, using a collision energy ramp defined between two lines from 50 to 5000 m/z: from 6/9 V (low mass, start/end) to up to 147/183 (high mass, start/end). MSMS scan time was set to 100 ms with an accumulated ion count 'TIC stop parameter' of 350 000 cps allowing a maximum accumulation time of 200 ms. After acquiring, masses from MSMS were excluded for 10 s. Wideband

enhancement was used to obtain a near-100% duty cycle on singly charged fragment ions. Lockspray of Glufibrinopeptide-B (m/z 785.8427) and leu-enkephalin (m/z 556.2771) was acquired at a scan frequency of 60 s. Data were lock-mass corrected post-acquisition.

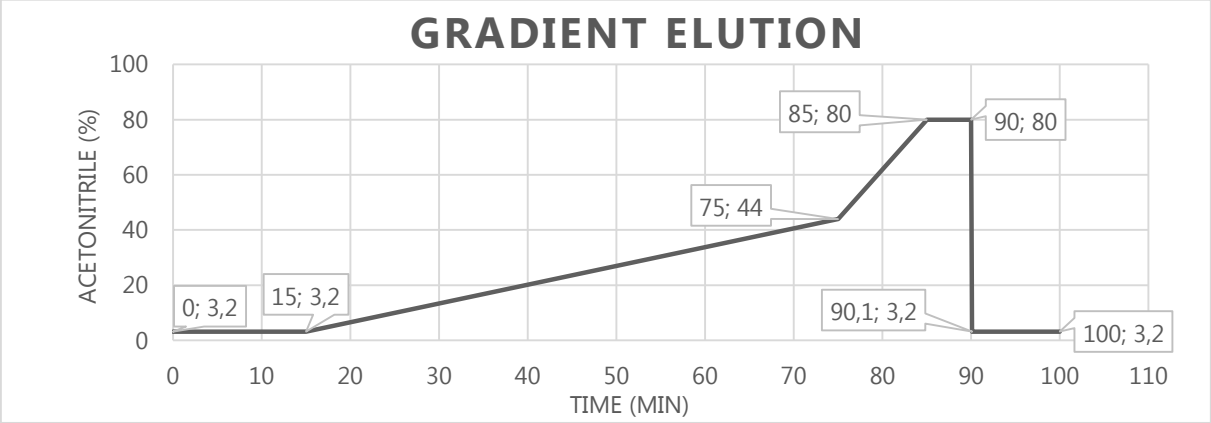


Figure 6.1 Gradient elution profile for the separation of propionylated histone peptides by RP-LC

LCMS for all other samples was performed by means of an Eksigent LC system coupled to a TripleTOF 5600. Peptides were separated using the gradient as shown in Figure 6.1. The mass spectrometer was applied with a ESI source in positive ion mode for DDA. Survey MS scans were acquired and up to 12 precursors exceeding the signal threshold (750 cps) were selected (m/z 400-1250).

Data analysis

The abundance of different peptide forms (desired, amidated at COOH, methylated at COOH, under- and overpropionylated) of seven peptides were monitored using their XIC's throughout all replicates. The conversion rate of each form was calculated as shown here for underpropionylation:

$$\boldsymbol{Amount\ of\ underpropionylation} = \frac{XIC\ underpropionylated}{XIC\ desired + XIC\ underpropionylated + XIC\ overpropionylated + XIC\ COOH - amidated + XIC\ COOH - methylated}$$

The area under the curve (AUC) itself was defined in two different manners, depending on the LCMS platform used. Both approaches are described into more detail below.

MS data acquired by the TripleTOF 5600 was processed as follows. Generated wiff-files were thoroughly searched using Protein Pilot 4.5 (Sciex), implementing the following special factors: purified histones, propionylation pre-digestion and propionylation post-digestion. Enzyme specificity was set to Arg-C and an FDR analysis was performed. The data were exported to Excel and filtered using 5% distinct peptide local FDR as cut-off value. The different peptide forms of eight peptides (underpropionylated, desired, overpropionylated) were isolated and XICs of these forms throughout all runs were made using

PiekView 1.2. To confirm correct annotation and localization of the propionyl groups, the same dataset was also searched using a Mascot 2.5 in house server (Matrix Science) and only spectra assigned to the same peptide by both ProteinPilot and Mascot were taken into account. Mascot parameters were the following: mass error tolerances for the precursor ions and its fragment ions were set at 15 ppm and 0.01 Da respectively, enzyme semi-specificity was set to Arg-C, allowing for up to two missed cleavage sites. Variable modifications included acetylation, dimethylation, trimethylation and propionylation on lysine (K), methylation and dimethylation on arginine (R) and oxidation of methionine. Lysine monomethylation was searched as the sum of propionylation and methylation since monomethylated lysine residues can still be propionylated. Propionylation was set as a variable modification at the N-terminus, serine (S), threonine (T) and tyrosine (Y). Only peptides with an expectancy value < 0.01 were considered correctly annotated.

Progenesis QI for Proteomics software and Mascot Deamon 2.5 were used to process raw MS data coming from Synapt G2Si and for further in-depth analysis of the samples. After alignment of the different runs to a single reference run (QC run, made from every sample of the experiment), raw data of all runs were processed using the default peak picking parameters with regard to sensitivity and chromatographic peak width. Features were filtered based on retention time and charge state (2+ to 5+) and the data were normalized to all features. Before exporting the data as an MGF-file, a top 5 MS/MS selection was taken to reduce the number of redundant spectra exported. These MGF files were searched using Mascot Daemon with following parameter set: decoy database, MS/MS tolerance of 0.3 Da, enzyme semi-specificity was set to Arg-C, allowing for up to two missed cleavage sites. Regarding modifications, the defined variables were: acetyl (K), amidated (C-term), amidated (DE), methyl (C-term), methyl (DE), methyl (K), methylpropionyl (K), propionyl (K), propionyl (N-term), propionyl (S), propionyl (T), and propionyl (Y). The data, obtained in Mascot, was filtered with an expectancy value of 0.01 and was exported as an XML-file and imported back into Progenesis QI software in order to link peptide identifications with peak areas. Then, the data was exported to Excel and the conversion rates were calculated as stated above.

6.3. Results

In the previous chapter several pitfalls in propionylating histones for bottom-up MS were disclosed: incomplete derivatization, aspecific propionylation and side reactions on carboxyl groups. During acquisition, considerable amounts of time are lost to these redundant peptide species and adding *in vitro* induced modifications in the search impedes identification of *in vivo* modifications. Equally, only relative abundance is robust enough not be affected by these modifications, making other methods such as direct inter-run relative quantification of AUCs imprecise. We thus set out to address these known issues by making adjustments to known protocols. An overview of the effects thereof is given in Table 6.1.

6.3.1. An attempt to tackle amidation of carboxyl groups in method A

As decribed in Chapter 5, propionylation method A suffers from unexpected amidation of carboxyl groups, next to the desired N-acylation at primary amines (Figure 6.2 A). This side reaction has not been described before and the exact nature of the reaction mechanism remains unclear. In Figure 6.2 B we make a suggestion of what this amidation mechanism may look like on a peptide C-terminus. In this representation, the carboxyl group itself becomes an anhydride, due to a side reaction between propionic anhydride and the peptide C-terminus. Hence, this reactive "mixed anhydride" can react whith ammonium hydroxide, when added to the reaction in a second step. As such, this will result in an amidated C-terminus and the formation of propionic acid. We tried to prevent this reaction from happening by changing the order in which reagentia were added in method A.reversed (method A.r) and method A.together (method A.t) or simply omitting ammonium hydroxide: method A.no NH_4OH (method A.n).

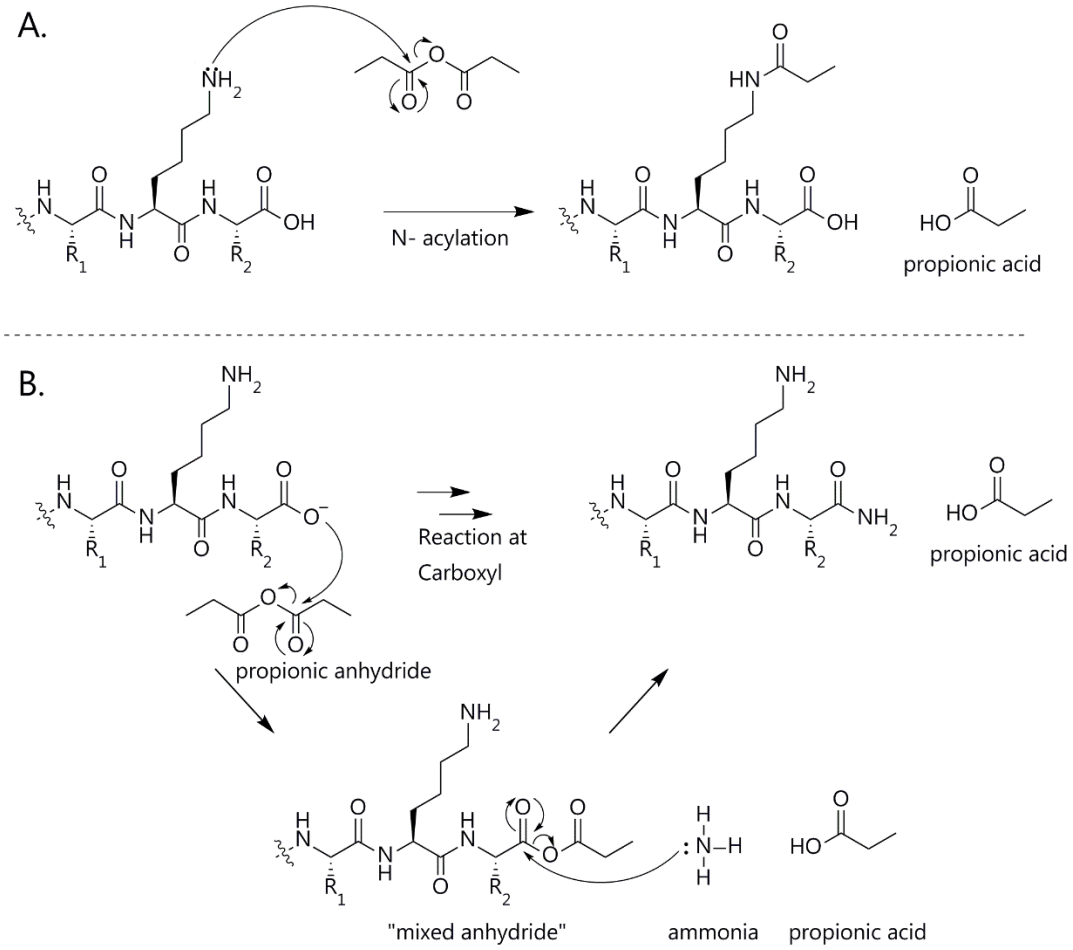


Figure 6.2 Amidation of carboxyl groups
A: Correct chemical derivatization of lysine containing peptides by propionic anhydride (K, Kme and N-termini). R1 and R2 represent amino acid side chains. After derivatization, the amine group on the unmodified K residue is modified with a propionyl group. Propionic acid is the side product of this reaction. **B**: Suggested reaction mechanism of the amidation of carboxyl groups (D, E and C-termini), which can occur as a side reaction during the chemical derivatization of histones by means of propionic anhydride and ammonium hydroxide. Hereby a mixed anhydride is formed, which reacts with ammonia and thereby generates an amidated carboxyl group.

6.3.1.1. Changing the order of adding reagents can turn amidation at COOH into underpropionylation – method A.r and A.t

In order to avoid amidation of carboxyl groups, we made some adjustments to method A and examined whether or not the amount of amidated peptides was reduced. We assumed (based on Figure 6.2) that by preventing the conversion of a carboxyl group in a very reactive "mixed anhydride", subsequent amidation would be circumvented as well. Therefore, we decided not to change the reagents used, only the order in which they were added to vacuum dried bovine histone. We reasoned that by mixing the same amount of ammonium hydroxide and bovine histones prior to (method A.r) or simultaneous with (method A.t) propionic anhydride, the latter would no longer react with carboxylgroups, since it will

readily react with the amines coming from the bovine histones as well as ammonium hydroxide to form correctly propionylated peptides and propionamide, respectively.

As anticipated, both customized methods no longer show any form of amidation at carboxyl groups. Unfortunately, besides hindering amidation, changing the order of adding reagents also affects efficient propionylation of primary amines on bovine histones, leading to a drastic increase in underpropionylated peptides. Just like amidation, this impaires identification as well as relative inter-run quantification of histone PTMs. We thus concluded that both methods are insufficient.

6.3.1.2. Omitting ammonium hydroxide can result in overpropionylation and turn amidation into methylation at COOH groups – method A.n

Knowing that nitrogen is needed as a nucleophile for the amidation of COOH groups, it seems reasonable not to add any nitrogen donors to the reaction when trying to avert amidation as a side reaction. Accordingly, we developed a method - based on method A - that no longer includes the addition of ammonium hydroxide: method A.no NH_4OH (method A.n). Only, if the reaction mechanism as suggested in Figure 6.2 is correct, leaving out the addition of ammonium hydroxide would result in the reactive "mixed-anhydride" intermediate as a final product. For this reason, water was added to further hydrolyze this anhydride into the original peptide (ending on a carboxyl group) and propionic acid. Taken into account our previous train of thought (Figure 6.2), the overall reaction mechanism of this propionylation method would be as depicted in Figure 6.3 A. However, the results did not confirm our assumption. Data analysis revealed another side reaction occurring: formation of methylesters on COOH groups and overpropionylation on S, T and Y. The former is probably caused by the use of methanol, which can react with the "mixed anhydride" intermediate (Figure 6.3 B). The hydroxylgroups of S, T and Y can also react with the excess of propionic anhydride, even though their reactivity is less compared to primary amines (Figure 6.3 C).

We thus conclude that omitting ammonium hydroxide in method A indeed stops amidation at carboxyl groups, but introduces other unwanted side reactions such as COOH-methylation and overpropionylation.

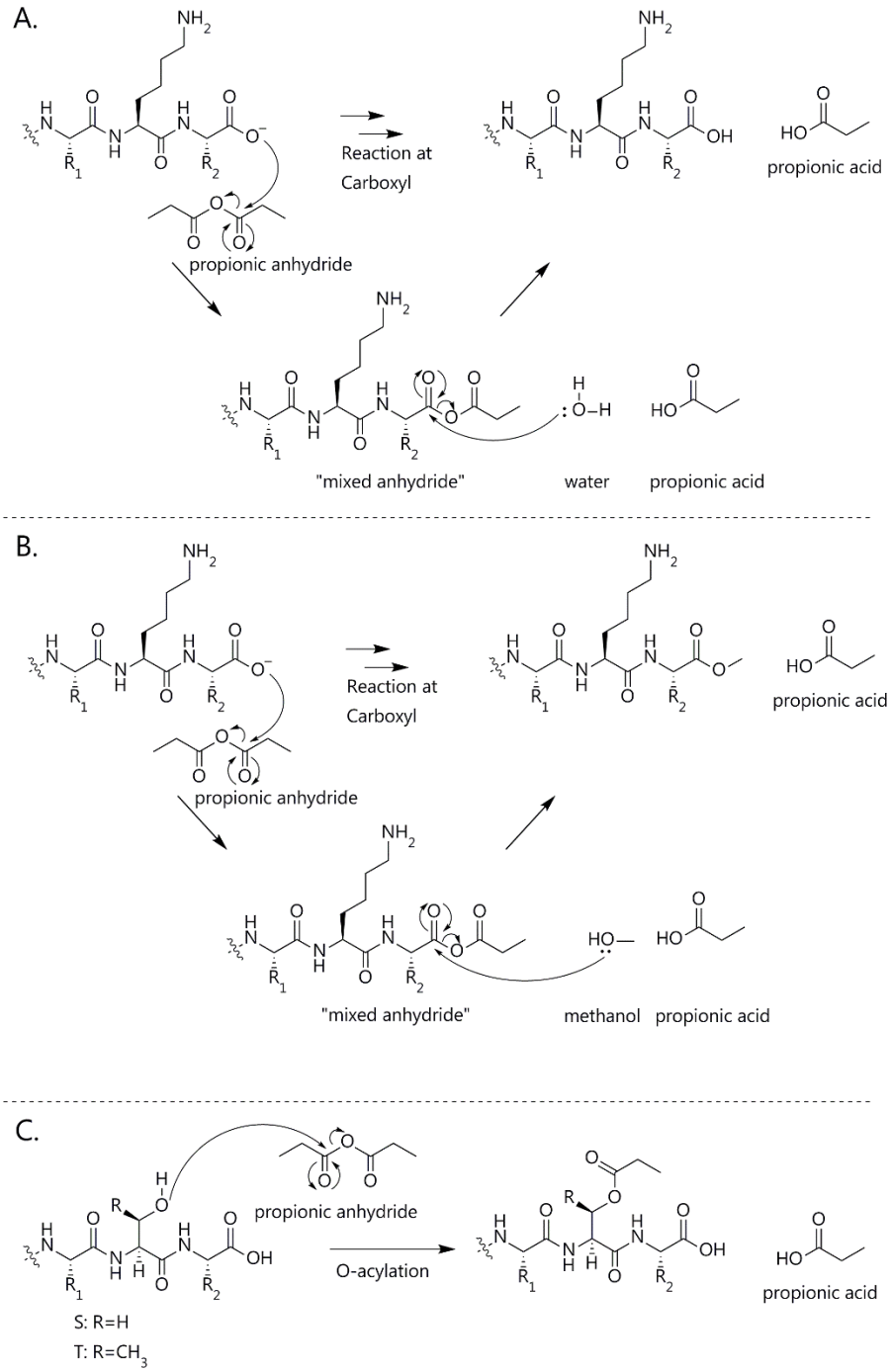


Figure 6.3 Hydrolysis and methylation of the "mixed anhydride"; aspecific propionylation

A: Addition of propionic anhydride to histone peptides can generate a "mixed anhydride" intermediate. This anhydride can be hydrolyzed by addition of water, thereby regenerating a the original COOH group. **B**: In the presence of methanol, hydrolysis as depicted in panel A is displaced by the reaction between the "mixed anhydride" and methanol itself. This results in the formation of methylesters at carboxyl groups of D, E and the C-termini. **C**: Aspecific propionylation can take place on the hydroxylgroups of S, T and Y by means of O-acylation. Panel C displays this reaction for both S and T residues. Y residues contain a phenyl group and are therefore less reactive, but they can react likewise.

Of interest, substituting methanol for isopropylalcohol in method A.n could stop the formation of methylesters on carboxyl groups. When we applied this, additionally no isopropylesters were detected. This was also described in the publication of Olcott et al., who studied the esterification of proteins by low molecular weight alcohols [214]. They report that primary alcohols other than methanol react more sluggishly with proteins, and ispropylacohol does not react appreciably. Overpropionylation on the other hand remained present when isopropanol was used. Moreover, we were surprised to see amidation recurring, even though no ammonium hydroxide was used in this method. This side product is potentially generated during digestion in ABC buffer.

6.3.2. An attempt to tackle inefficient propionylation of primary amines in method B and C – method B.s

Since amidation of carboxyl groups is a new phenomenon with an unknown reaction mechanism, it is difficult to control this side reaction. Hence, we equally set out to further improve a known propionylation protocol that does not suffer from amidation. As described in Chapter 5, both methods based on the protocol of Garcia et al. [128] (method B and C) were hindered by underpropionylation. In order to optimize the reaction efficiency, we developed method B.separate (method B.s), where two major adjustments were made compared to methods B and C (Chapter 5). First, ammonium hydroxide is now added as a final step during propionylation, in stead of prior to propionic anhydride. The reason therefore is that the large amount of primary amines coming from ammonium hydroxide could compete with the primary amines on the bovine histones themselves, hindering proper derivatization. Also, methanol and propionic anhydride were no longer mixed prior to the propionylation reaction, but instead were added separately. The main reason therefore is that propionic anhydride will readily react with methanol, once the propionylation reagent is made. This way, a lot of the reactivity of this mixture might be lost, even before it is added to the histone sample. Note that also other research groups acknowledge this to be an issue, as they recommend to prepare fresh reagent every three samples and equally stress the importance of working in a streamlined manner, without any interruptions [128,179].

As we reached our objective of limiting the amount of underpropionylation (on average < 2%), other side reactions emerged instead. We now found overpropionylation (up to 18%) and amidation (up to 31%) to be the main problems. The latter can be due to the presence of ABC and ammonium hydroxide in the reaction. Even though methanol was used, no significant amount of methylesterification was detected. In conclusion, we were able to improve propionylation efficiency on primary amines by making some small adjustments to method B, but this was at the cost of a loss in specificity. Besides, amidation of carboxyl groups appeared as an unwanted side reaction.

6.3.3. An attempt to improve specificity and efficiency by omitting or changing the buffering system

Since avoiding amidation at COOH groups, as well as preventing underpropionylation proved to result in other side reactions, we tried some new approaches for the derivatization of histones. Method E is based on literature, methods F till H are new, according to our knowledge.

6.3.3.1. Propionylation without buffering system – method E

In a first attempt, histones were propionylated by use of a protocol mentioned in a publication of Smith et al. [130]. This method (method E) includes no buffers (such as ABC or ammonium hydroxide) that are typically used to maintain a pH between 8 and 10. The only reagent used, next to propionic anhydride, is propionic acid. For a detailed description, see section Methods (section 6.2).

Even though the primary amines on the histones were all adequately propionylated, this derivatization method was far from specific. A large amount of the peptides appeared in their overpropionylated form (up to 89%).

6.3.3.2. Propionylation with ethanolamine, triethanolamine or TEAB – method F, G and H

Because the use of ammoniumhydroxide or ABC was associated with amidation at carboxyl groups and the use of methanol could lead to the generation of carboxymethylesters we decided to test whether or not omitting methanol and at the same time using another buffering system could prevent these side reactions from happening. We investigated the use of ethanolamine (pKa = 9.5), triethanolamine (pKa = 7.74) and TEAB (pKa = 10.75) (Figure 6.4). The methods (methods F, G and H respectively) are described in detail in section 6.2.

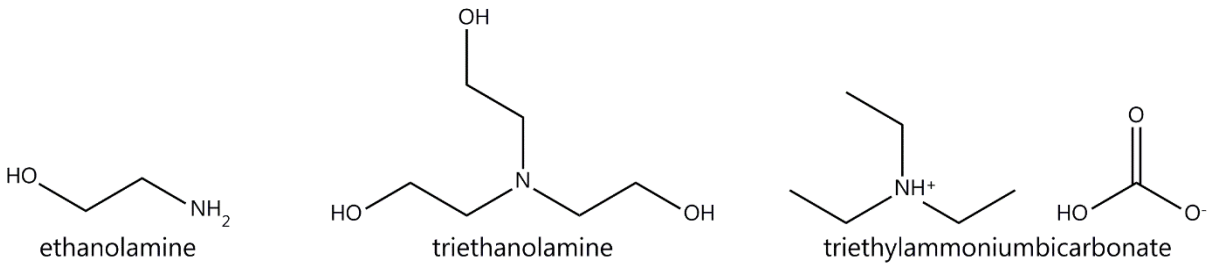


Figure 6.4 Chemical structure of ethanolamine, triethanolamine and TEAB

Indeed, no amidation nor formation of methylesters could be detected on carboxylgroups for either of these protocols. Also, the reaction was efficient on primary amines as no underpropionylated peptides were identified. Nevertheless, all three methods struggle with a lack of specificity as each of them shows

a lot of overpropionylation: 43% (method F), 61% (method G) and 62% (method H) on average. As such, these methods do not suffice for the propionylation of histones.

6.3.4. **Dealing with aspecific overpropionylation by preventing or reversing it**

Foregoing methods illustrate that avoiding inefficient propionylation (underpropionylation), often goes at the cost of introducing aspecific overpropionylation. Nevertheless, some propionylation strategies (method D, E, F, G and H) only have overpropionylation as an unwanted side effect and could thus result in a proper method if this issue is addressed. Therefore, three different strategies were performed on aforementioned protocol H, which – based on our analysis - suffered from no other issues than overpropionylation. **(i)** At first, we tried to prevent overpropionylation by lowering the molar excess of propionic anhydride, compared to the average amount of primary amines on bovine histones. We thus lowered the volume of propionic anhydride added to the reaction. Next, we verified two different approaches for reversing overpropionylation, after the final propionylation reaction post digestion: **(ii)** boiling of propionylated peptides during 1 h and **(iii)** adding hydroxylamine.

6.3.4.1. **Lowering propionic anhydride concentration down to a 5-fold molar excess**

One approach to circumvent aspecific propionylation is lowering the concentration of propionic anhydride from the estimated 420-fold molar excess used in method H to a 42-fold (method H 42x), 20-fold (method H 20x) and 5-fold (method H 5x) molar excess (per primary amine group). Also, propionylation of the samples was only performed once prior to and post digestion instead of twice. This drastic decrease is based on literature where the molar excess of anhydride reagents ranges from 5x over 200x up to 1000x [215–219].

As can be seen in Figure 6.5, reducing the surplus of propionic anhydride decreases the amount of overpropionylated peptides. While the portion of peptides with propionyl groups at S, T and Y is still high for a 42x molar excess of propionic anhydride, there is a significant decrease for a 20x surplus and overpropionylation is almost reduced to zero for 5x molar excess. Nevertheless, these methods still do not meet expectations as they simultaneously display an increase in underpropionylated peptides.

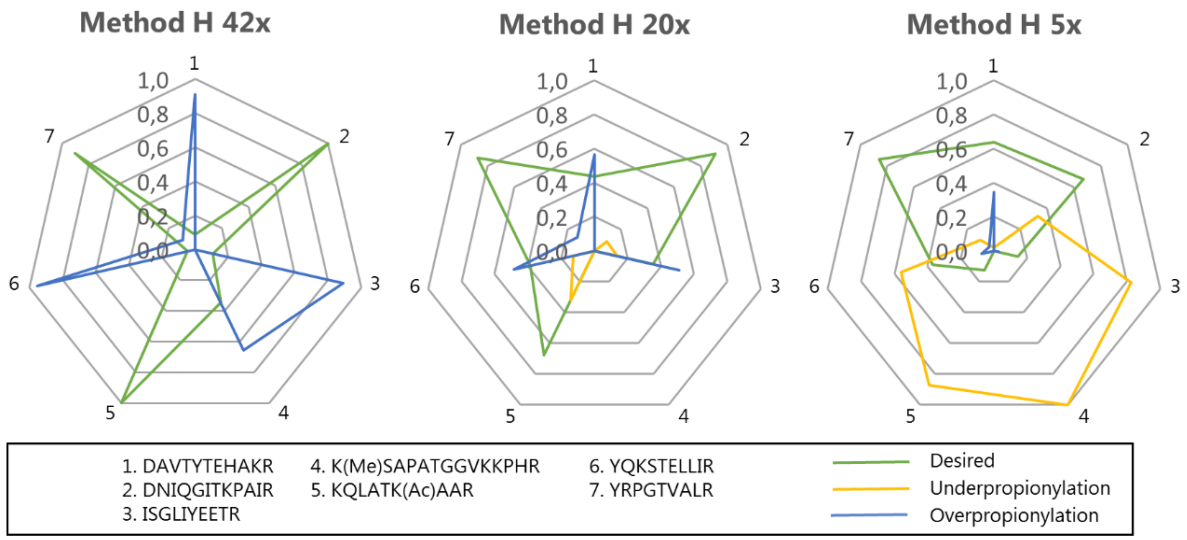


Figure 6.5 Lowering the molar excess (42x, 20x and 5x) of propionic anhydride per primary amine prevents most overpropionylation but increases underpropionylation
Each method is represented by a radar chart showing the average contribution of over-, under- and desired propionylation for 7 peptides monitored. Each peptide is located on one angle of the radar chart and each peptide form is represented by another color. The abundance of a specific peptide form is shown on the radius, whereby a conversion rate of 0 is located in the center, increasing outwards.

In conclusion, we can state that lowering the molar excess of propionic anhydride per primary amine helps in preventing overpropionylation but at the same time hinders efficient propionylation on the targeted primary amines.

6.3.4.2. **Reversing overpropionylation through boiling or adding hydroxylamine**

Given the latter conclusion, we set out to reverse this overpropionylation, instead of avoiding it. Based on literature we found two strategies: **(i)** boiling of propionylated peptides during 1h (method H42x boil) [220] and **(ii)** adding hydroxylamine (method H 42x ha) [221–223], and implemented these on method H 42x since overpropionylation is the only known side reaction for this approach. The reaction mechanism of the latter is displayed in Figure 6.6.

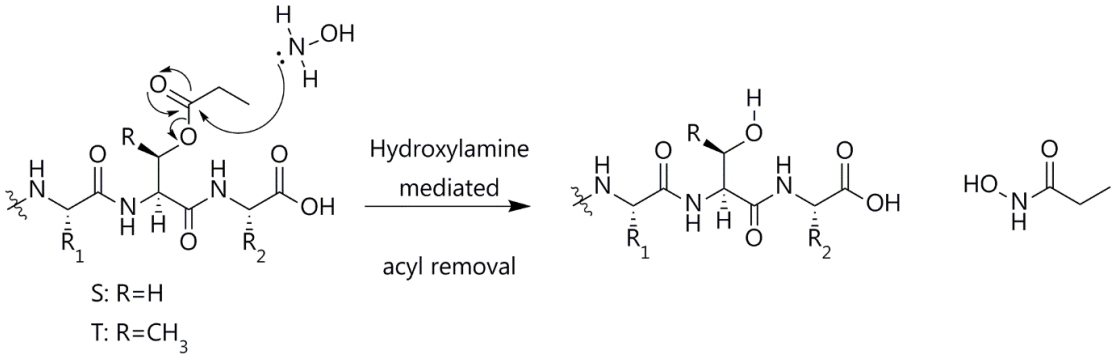


Figure 6.6 Reversing overpropionylation by means of hydroxylamine

As illustrated by Figure 6.7, both boiling and the addition of hydroxylamine result in the reversing overpropionylation, without a substantial increase in underpropionylated peptides. Boiling the samples decreases aspecific propionylation of all 7 peptides monitored, but it appears to be less efficient for some peptides, such as DAVTYTEHAKR. Treatment with hydroxylamine on the other hand was very effective with a negligible discrepancy for peptide sequence. The average conversion into correctly propionylated peptides increased up to 95 %. The only disadvantage of using this method is the loss of acetylation at S, T and Y residues as a biological modification. This PTM will most probably be removed simultaneously with the propionylgroups and can thus no longer be identified, nor quantified. However, this modification is very rare compared to other PTMs that can now be characterized more reliable. Additionally, it is a very challenging modification to be annotated correctly by MS, as it is extremely difficult to distinguish it from neighboring and more abundant K acetylations.

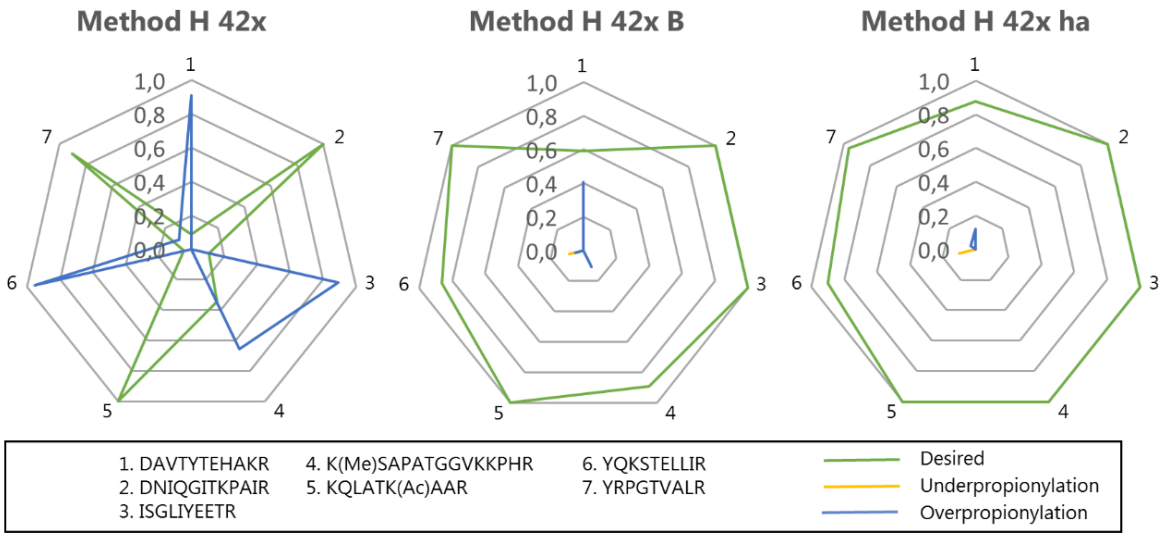


Figure 6.7 Overpropionylation can be (partially) reversed by boiling the sample for 1h or adding hydroxylamine

Each method is represented by a radar chart showing the average contribution of over- (blue), under- (yellow) and desired propionylation for 7 peptides monitored. Each peptide is located on one angle of the radar chart and each peptide form is represented by another color. The abundance of a specific peptide form is shown on the radius, whereby a conversion rate of 0 is located in the center, increasing outwards.

Thus, both methods seem promising as a new strategy to reverse overpropionylation without causing underpropionylation. At this point, adding hydroxylamine is the method of choice as it can generate on average 95% of correctly propionylated peptides.

6.4. Discussion

Histone analysis by bottom-up MS requires an additional step during sample preparation as compared to the traditional proteomics strategy: chemical derivatization by acylation reaction at amine groups. This is mainly because histones are highly enriched in basic Aa residues (K and R), which comes as an advantage for the proper binding to DNA but results in generation of very short peptides for LCMS analysis after trypsin digestion. Therefore, histones are mostly derivatized prior to and after trypsin-mediated cleavage by means of propionylation of accessible amine groups. However, the technical framework of this propionylation step should not induce additional variation that obscures biological changes.

In Chapter 5 we performed an extensive investigation of side reactions for 4 commonly used propionylation protocols (method A till D), using DDA and peptide quantification performed via XIC after precursor alignment. This analysis disclosed several pitfalls: side reactions on carboxyl groups (amidation, methylation), incomplete derivatization and aspecific propionylation at S, T and Y residues. These pitfalls hinder identification and impair direct comparison of precursor intensities of biological modified peptides. However, relative abundance proved to be a very robust relative measurement, that quantified PTM very similarly independent of the propionylation protocol applied. This implies that research groups using one or the other protocol are likely to provide equally results when using this measurement to quantify changes between different biological samples [224]. Note that correction factors for ionization efficiency are required when aiming at quantifying the true relative abundance in a single sample and that the robustness here applies only to the comparison of different relative abundance values. Nevertheless, in order not to generate large amounts of uninformative new precursors that are being selected for MSMS during DDA and at the same time enable "inter-run relative quantification", we set out to address these pitfalls in propionylation and develop a protocol that only leads to correctly propionylated peptides. The results are summarized in Table 6.1.

First, we tried to avoid amidation of carboxyl groups by making some adjustments to method A. In method A.r and A.t, we mixed the same amount of ammonium hydroxide prior to or simultaneous with propionic anhydride, which no longer led to the generation of amidated carboxyl groups. However, due to increase in underpropionylated peptides we refrained from further developing this strategy. Leaving out ammonium hydroxide as such, as done in method A.n, could also prevent amidation of carboxyl groups, but on the other hand induced methylation of these groups and led to aspecific propionylation at S, T and Y residues. The former is probably due to an O-acylation reaction between the very reactive "mixed anhydride" and methanol. This reaction did not occur in the presence of ammonium hydroxide

(method A, A.r and A.t) because amine groups are more reactive than hydroxyl groups. In conclusion, the adjustments made to method A failed to provide an optimal propionylation method.

Next, we tried to convert the underpropionylation in method B and C. Despite the fact that all underpropionylation was countered by omitting the mixing of methanol and propionic anhydride prior to the reaction, as well as by adding the ammonium hydroxide after the propionylation reagent, new pitfalls (amidation and overpropionylation) shut the door on this protocol.

Subsequent, we investigated the effect of interfering with the buffer system, either by using no buffer at all or changing the buffer system (ethanolamine, triethanolamine and TEAB): method E till H. Unfortunately, all methods led to an increase in overpropionylation. The reason therefore probably is the reactivity of an anhydride towards hydroxylgroups. While primary amines are more reactive, acylation of hydroxyl groups becomes favorable once these are all propionylated. Nevertheless, if overpropionylation is the only pitfall and these methods are not hindered by any other unwanted reactions, it is worth trying to address this issue specifically.

Hence, we tried to tackle this aspecific propionylation in two ways: **(i)** preventing it from happening by lowering the molar excess of propionic anhydride or **(ii)** reversing it by means of boiling or the addition of hydroxylamine. The first approach was partially successfull given that overpropionylation decreased in correlation to the reduction in propionic anhydride added to the samples. Yet, underpropionylation started to increase simultaneously, and finding the optimal molar excess to balance between both pitfalls is difficult. Furthermore, researchers often do not know the exact amount of sample – let alone primary amines- they have on their hands, which would be required for converting an ideal molar excess into a specific volume of propionic anhydride. Taking this into account, reversing propionylation of hydroxylgroups on S, T and Y might be a better strategy. This goal could be achieved by both boiling and by the addition of hydroxylamine. Heating the sample was able to break the ester bond on most peptides but unfortunately the amount of overpropionylated peptides remained too high for some peptides, such as DAVTYTEHAKR where 41% of the peptides were left overpropionylated. The hydroxylamine mediated acyl removal on the other hand was more efficient with an average conversion rate of 95%. The addition of this very reactive nucleophile can reverse overpropionylation, without removing propionylation on primary amines. Hence, all protocols generating only overpropionylated products as unwanted side products can be turned into good propionylation protocols by addition of hydroxylamine post propionylation. Yet, when using a method for reversing ovepropionylation, one should always take into account that acetylgroups on S and T residues will no longer be identified.

Table 6.1 Customizing a method in order to tackle known pitfalls
In order to address the known pitfalls, methods were customized. The effect of these adjustments on the average contribution of each peptide form was monitored. This contribution could be none (/), increase (↑), decrease (↓) or the adjustment could lead to complete reduction (↡) of the peptide form.

Based on - ***pitfall***	**Method name**	**Adjustments made**	**Pitfall**			
			Amid	Meth	Und	Over
Method A ***- Amid***	A.r	Reverse order of adding NH_4OH and PA reagent	↡	/	↑	/
	A.t	Add NH_4OH and PA reagent together	↡	/	↑	/
	A.n	Do not add any NH_4OH	↡	↑	/	↑
Method B and C ***- Under***	B.s	Do not mix methanol and propionic anhydride in advance Add NH_4OH at the end	↑	/	↡	↑
NEW	E	No buffer	/	/	/	↑
	F	Ethanolamine buffer	↑	/	/	↑
	G	Triethanolamine buffer	/	/	/	↑
	H	TEAB buffer	/	/	/	↑
Method H - ***Over***	H 42x	42x molar excess of propionic anhydride	/	/	/	↓
	H 20x	20x molar excess of propionic anhydride	/	/	↑	↓
	H 5x	5x molar excess of propionic anhydride	/	/	↑	↡
Method H 42x ***- Over***	H 42x boil	Boil sample for 1h after propionylation	/	/	/	↓
	H 42x ha	Reaction with hydroxylamine after propionylation	/	/	/	↡

In conclusion, we were able to develop a protocol (method H 42x ha) that attains efficient propionylation on primary amines, has no remaining aspecific propionylation of S, T and Y residues and displays no side reactions such as amidation or methylation of carboxyl groups. Despite the loss of information on *in vivo* acetylated S and T residus, this propionylation protocol has several advantages over other methods. First of all, sensitivity will rise as the MS signal of the same peptide is no longer spread over different MS precursors, generated by unwanted side reactions or inefficient propionylation. Additionally, no precious acquisition time will be lost by generating MSMS spectra of these differentially *in vitro* modified peptides instead of spectra belonging to peptides with the interesting *in vivo* modifications. But not only data acquisition will benefit from using both efficient and specific propionylation protocol: also the confidence of peptide identifications will increase, considering that less variable modifications such as propionylation on S, T and Y have to be taken into account when performing a database search. Alternatively, more *in vivo* modifications can be permitted before causing a combinatorial explosion. This in turn will lead to more biologically relevant conclusions. Finally, sample preparation will require less time as only one round of propionylation is required prior to as well as post digestion.

"A scientist is not a person who gives the right answers, he's the one who asks the right questions."

— Claude Lévi-Strauss

CHAPTER 7:

FINAL DISCUSSION AND GENERAL CONCLUSIONS

Partially based on "**Histone proteolysis: A proposal for categorization into 'Clipping' and 'Degradation'**"

Maarten Dhaenens, Pieter Glibert, Paulien Meert, Liesbeth Vossaert and Dieter Deforce

BioEssays (2015), 37(1), 70-9

7. FINAL DISCUSSION AND GENERAL CONCLUSIONS

This thesis aimed at studying histone PTMs from both a biological and technical point of view. Throughout the first part, we report histone H3 cleavage, mediated by a serine protease, as a new PTM in hESC and investigate its occurrence in a biological context. In the second part we further elaborate on the technical pitfalls associated with uncovering such histone PTMs by bottom-up MS.

7.1. Histone clipping: the pursuit of an epigenetic outcast

7.1.1. Oct4 expression is not directly related to H3 clipping activity

In the first part of this dissertation we focused on the N-terminal truncation of histone H3 during early hESC differentiation. A similar event was reported earlier in 2008 by Duncan et al. in mESC [184]. This manuscript strongly implies a cause-and-effect relationship between H3 clipping and differentiation. In order to disclose a similar link between both this clipping event and and the expression of the pluripotency marker Oct4, we set out to monitor the hESC over a longer period of time (14 days) and subjected them to three different types of differentiation: **(i)** stem cells were maintained in their undifferentiated status by culturing the cells in E8 medium, **(ii)** spontaneous differentiation was induced by omitting bFGF from the culture medium and **(iii)** addition of RA directed enforced differentiation. Despite the clear decline in Oct4 expression, no change in H3 proteolysis activity could be observed. Even though these results seem to be conflicting with the findings of Duncan et al. – who claim that H3 clipping to be differentiation-related - a thorough analysis of their results leads us to believe otherwise. In one of the cathepsin L inhibition experiments, the authors monitored H3 clipping together with Oct4 expression. While the Oct4 expression kept on following the same trend, with or without cathepsin L inhibition, the clipping event was clearly influenced by this inhibition. Thus, similar Oct4 expression, but nevertheless H3 cleavage was affected [184]. Still, our findings seem to suggest that cells require a certain degree of multipotency for the clipping to occur. This was illustrated by the fact that terminally differentiated Raji cells are incapable of H3 clipping in contrast to multipotent THP-I cells. A possible explanation for this is the more loosened chromatin structure associated with pluripotent cells, which increases the susceptibility for histone clipping. However, more cell types need to be investigated to exclude the possibility that this finding simply comes down to a difference in enzyme expression between different cell lineages.

7.1.2. The truncation of H3 is not cell cycle regulated

Next, the undulating pattern of cH3 seen when hESC are cultured on MEFs (Figure 4.9), together with the known effect of RA addition on the cell cycle itself [225], drove us to hypothesis that the event instead is cell cycle-dependent. Duncan and co-workers considered this theory as well, based on flow

cytometry data that revealed a prolonged S phase of the mESC upon differentiation. Moreover, the preferred clipping of the H3.2 isoform, whose expression is tightly replication-coupled, might indicate a preference of histone proteolysis towards the S phase. However, synchronized cells showed no change in H3 clipping in neither of the cell cycle phases. We thus conclude there is no direct link between the cell cycle and histone clipping.

7.1.3. **Culture conditions alter H3 clipping pattern in differentiating hESC**

To further extend the context of histone clipping, we investigated two different hESC cell lines under two different culturing conditions. Surprisingly, the latter variable had a major impact on the H3 clipping pattern in differentiating hESC. While both cell lines showed continuous clipping under feeder-free culturing conditions, this no longer applies to hESC cultured on a MEF feeder layer. In this setting, hESC cleave H3 in an identical temporary window compared to mESC: the clipping process starts on day 2 after RA addition and reaches a maximum at day 4. Henceforth, the cH3 intensity fades towards the end of the experiment. Even though Duncan et al. do not specifically mention the effect of culturing conditions on the cleavage pattern, they saw a similar phenomenon during EB formation, where the cH3 band appeared early and remained until 14 days after the start of differentiation. Nevertheless, when monolayer differentiation was induced by means of RA, H3 proteolysis appeared to be temporary, arising at day 2 and 3 and disappearing thereafter [184]. Only very recently the group of Jensen also noted that culture conditions can impact histone proteolysis [226]. Histone clipping was induced during the cultivation of human hepatocarcinoma cell line in 3D culture but not in 2D culture. The dissimilarity in proteolytic activity between culture conditions for both this cell line and hESC might be explained by the difference in composition, in hESC colony architecture, in cell signaling, and in the biomechanics of the cellular microenvironment [227]. In general, these features act in concert to provide the necessary cues, regulating cell function in both developing and adult organisms. Substrate-cell, as well as cell-cell interactions can activate specific mechanotransduction pathways that are involved in the regulation of the stem cell fate. Mechanical factors and geometrical properties can thus have significant influence on regulating stem cell activities [228]. Also histone H3 clipping might one of those affected processes. Moreover, one can doubt whether the 2D culturing system for hESC (MEF feeder layer or feeder free) even comes close to representing the 3D *in vivo* conditions, rendering the biological relevance of processes such as histone clipping less trustworthy. Additionally, to eliminate the possibility that this clipping event is actually a technical *in vitro* artefact, we included several controles: adding a spike-in of **(i)** biotinylated and **(ii)** non-biotinylated calf H3 during extraction, **(iii)** whether or not omitting protease inhibitors at the time of sample preparation and **(iv)** the application of direct boiling after cell harvesting.

Each of these measures confirmed our assumption that H3 proteolysis is not merely an *in vitro* side effect, but takes place in the nucleus itself prior to extraction.

7.1.4. A broader context: categorizing histone proteolysis into histone clipping and histone degradation

When fitting these results into a broader context one can remark that also other research groups have linked histone disintegration and developmental processes. However, the autors often do not state whether complete degradation or specific clipping was studied. Therefore, we have categorized earlier reports linking histone proteolysis and development into two different functional classes: (complete) histone degradation on one hand and the epigenetically connoted histone clipping on the other.

First, the following events illustrate that overall histone degradation occurs predominantly during important developmental transitions. **(i)** Both through enzymatic degradation and mediated by the proteasome histones are degraded during spermatogenesis [229]. **(ii)** These sperm histones (protamines) then have to be degraded once again during early embryogenesis. The enzyme responsible for this was recently found to be cathepsin L. **(iii)** Also earlier eukaryotes go through complete histone degradation during specific stages of development. For example, H3 is degraded during *Tetrahymena* macronucleus formation, with H3A21 being the most abundant fragment [230]. **(iv)** Some intracellular pathogens such as Chlamydia seem to be able to selectively degrade their own histones upon infection [231]. Taken together, developmental histone degradation can be expected to be an omnipresent phenomenon in eukaryotic organisms, especially during reproduction. However, the relative lack of studies focusing on the mechanisms underlying these histone turnover events also conceals its relation to histone clipping.

More specifically, next to these events of (complete) histone degradation, H3-specific clipping has been raised as being of potential epigenetic importance in various biological processes. Here, we report these findings in a chronological order (summarized in Table 7.1). **(i)** In 1979, Allis et al. demonstrated that the micronuclei of *Tetrahymena thermophila* contain two electrophoretically distinct forms of histone H3. The faster species is derived from H3 by a proteolytic cleavage which removes six residues from the amino terminus [192,193,232]. **(ii)** Falk and co-workers were among the first to report the occurrence of histone H3 clipping in infection and dissease when they showed that the FMDV expresses the so-called protease 3C in host cells, which mediates clipping at H3L20 [187,188,233,234]. **(iii)** As mentioned earlier, the N-terminus of H3 is clipped from A21 to K27 during mESC differentiation [184]. **(iv)** Almost simultaneously with the publication of H3 trunctation in differentiating mESC in 2008, Santos-Rosa et al. reported a protease activity in yeast that was induced under conditions of nutrient deprivation and sporulation [191]. Five years later, Xue et al. appointed PRB1 as an enzyme responsible for this event

[205]. Similar to cathepsin L activity, the yeast endopeptidase cleavage of H3 occurred in a PTM-dependent matter. The authors showed that H3 truncation precedes H3 eviction from induced promotors, by removing repressive marks. **(v)** Moreover, glutamate dehydrogenase was discovered to also have H3 protease specificity in chicken liver [195]. Given that the truncated H3 form was only observed in adult chicken livers and remained absent in young chickens, a role for H3 clipping in aging has been suggested [194,235]. Within the same context, Kanungo and co-workers reported earlier the H3 truncation in aging quail liver [190,236]. Yet, they have not been able to pinpoint the responsible enzyme, nor could they unravel the epigenetic mechanism involved. **(vi)** Recently, histone H3.1 was demonstrated to be a potential legumain substrate in colorectal cancer, but the *in vivo* functional implications of this legumain nuclear activity are yet to be explored [237]. **(vii)** In 2014, a study of the post-lactational regression of the mammary gland revealed that PTMs drive cathepsin D into the nucleus to cleave H3. One year later, Arnandis et al. revealed enzyme redundancy when they showed that also Calpain-1 was internalized within nuclei and found to be present in the nuclear chromatin-enriched fraction, associated with histone H3. This developmental regression stage is an intriguing biological phenomenon, and its further elucidation is of great relevance to breast cancer research [238].

Based on this overview of histone clipping events, we can agree with the notion that H3 clipping could very well represent a common feature during developmental processes, such as differentiation and aging. Nevertheless, since the link between histone proteolysis and epigenetics only started to emerge during the last decade, conclusive evidence is still lacking. Uncoupeling controlled histone clipping from continuous histone degradation will probably prove to be the biggest challenge.

7.1.5. **N-terminal cleavage sites are targeted by a serine protease**

If we acknowledge H3 clipping during early hESC differentiation as a potential epigenetic mark, two key questions remain. First, where does this clipping take place in the chain of AA? Second, which enzyme mediates this event?

The former was addressed by implementing complementary techniques. Western blotting enabled to confirm the N-terminal nature of cH3 and could moreover pinpoint A21 as a first cleavage site. MS based analysis revealed residue 31 as an additional location for proteolysis. Finally, R26 was appointed by means of edman degradation, at least for the H3.1 and/or H3.2 isoforms.

For the identification of the enzyme mediating this N-terminal H3 truncation, we applied a targeted as well as a general approach, using both specific inhibitors for cathepsin L and general inhibitors for a specific enzyme class. **(i)** We assumed cathepsin L to be a major candidate as this enzyme is responsible for the clipping in mESC and has been proven to be able to clip H3 *in vitro*. Nonetheless, none of the

specific cathepsin L inhibitors nor the inhibitors for cysteine proteases in general could counter the truncation of purified calf H3 when incubated with nuclear extract (NE) that already had been proved to contain clipping ability. We thus conclude that cathepsin L is not the main H3 clipping mediator in hESC, even though Duncan et al. have proven otherwise in mESC. At least this finding is rather remarkable, given that both the biological context (stem cell differentiation) and cleavage sites are very similar. **(ii)** In analogy, we also subjected the *in vitro* H3 clipping capacity of the NE to several broad spectrum protease inhibitors (serine proteases, cysteine proteases, metalloproteases, aminopeptidases and/or aspartyl proteases). Out of all these inhibitors only AEBSF, a serine protease inhibitor, was able to intervene with H3 proteolysis. Hence, the proteolytic activity in hESC seems to be carried out by a serine protease.

Since H3 cleavage is extensively studied as an epigenetic template in general, we compare our findings concerning the cleavage site and the enzyme involved to previously published research. As displayed in Table 7.1, up until now all H3 clipping reported was located in the N-terminus. This is in contrast with the truncation of H2A, that occurs at the C-terminus in e.g. chronic lymhocytic leukemia and the myeloid THP-I cell line [171]. The specific localization of the H3 cleavage site itself differs, depending on the biological context involved. The same can be said for the enzyme mediating the clipping event. Moreover, only during the past decade convincing evidence was found for some of these enzymes concerning their ability to migrate to the nucleus where they maintain enzymatic activity [195,205,239–244]. This forces researchers to keep an open mind towards alternative functions of known proteins [245]. Both cathepsin L and cathepsin D were intitially described in cytoplasmic vesicles as proteinases that degrade protein substrates with broad specificity [246,247]. PRB1 in yeast also has yet been associated with vacuolar degradation for a long time [248], and GDH is mainly known as a metabolic enzyme.

Table 7.1 Overview of the known H3 clipping events: organism, biological context, cleavage site, enzyme and protease class

Organism/tissue	Biological context	Cleavage site	Enzyme	Protease class
S. Cerevisiae	Sporulation or starvation	A21-T22		Serine [191]
		K23-A24	PRB1 [205]	
T. thermophila/ micronuclei [192,193,232]	Demethylation mechanism	T6-A7	/	/
Neisseria meningitidis [249]	Virulence factors	/	App and MspA	Serine
FMDV/infected BHK cells [188]	Host cell transcription	L20-A21	Protease 3C	Serine
Chicken/liver [194,195]	Aging	K23-A24 K27-S28	Glutamate dehydrogenase	Cysteine
Mouse/ESC [184]	Differentiation	A21-T22 T22-K23 K23-A24 A24-A25 R26-K27 K27-S28	Cathepsin L	Cysteine
Mouse/ mammary glands	Post lactational regression	K23-A24	Cathepsin D [238]	Aspartyl
		K9-S10	Calpain 1 [250]	Cysteine
Human/tumor [237]	Colorectal cancer	/	Legumain	Cysteine
Human/ ESC	Differentiation	A21-T22 R26-K27 A/S 31-T32	/	Serine

7.1.6. **Histone proteolysis: an epigenetic outcast**

Histone proteolysis has a long history of disregard. The implications thereof on scientific research are twofold. First, many more proteolysis events have most probably been encountered in the past, but remained outside the scope of the authors at that time. One very remarkable recent case shows additional histone H3 bands in HIV latency infected cell lines [251]. Besides, when looking back at the publications concerning histone proteolysis over the last 50 years, it has always been treated in a stephmotherly fashion by the scientific community. Although the first papers started to suggest that histone truncation might greatly impact transcription in the 70s and 80s, the interest in this biological event seemed to completely fade away during the 90s and early 21st century. Nevertheless, 10 years later, a second wave if interest seemed to be upon us with the publication of the first evidence of epigenetic potential for histone clipping in mouse and yeast. Surprisingly however, this promise was again not fulfilled and histone clipping is still not picked up as a relevant PTM by the broader scientific community, at a time where over 5000 papers are published on histones every year.

Next to the immediate effect of brushing aside histone clipping as a potentially important process in biology, the disregard of this event can also have indirect effects on experiments aiming other processes involving histones. Any experimental approach that targets histone tails and their modifications is prone to the effects of both *in vivo* and *in vitro* histone proteolysis. Epigenetic screening techniques such as the detection of modifications by means of specific antibodies or the immunoprecipitation of modified tails are entirely blind to these effects. Considering the intensity of the cleaved fragment of H3 in differentiating ESC from mouse and human, it is fair to state that up to half of the whole histone H3 content can be clipped at certain time points in ESC differentiation [184,252]. It is thus definitely not a secondary phenomenon, whose impact can be treated with disregard. According to our knowledge however, researchers have rarely taken specific care to avoid this technical pitfall while studying histones.

7.2. Properly reading the histone code through bottom-up MS

MS analysis of histone PTMs is typically carried out by a bottom-up approach. In this workflow, histones are often chemically derivatized on K residues prior to and post digestion to make sure enzymatic cleavage occurs only at R residues and thus yield sufficient usable peptides. The specificity and effiency of this procedure will have an influence on different aspects of the histone analysis pipeline. Therefore it is crucial to fully understand the nature of this chemical derivatization prior to solving biological questions. We investigated different approaches for propionylation and examined the efficiency on primary amines (underpropionylation), the specificity of the reaction (overpropionylation) as well as the occurrence of unwanted side reactions such as methylation and amidation at carboxylgroups. In the following paragraphs, we will discuss the different pitfalls that emerged, their implications on data acquisition and data analysis, and elaborate on the efforts made in trying to tackle these problems. A summary thereof is given in Figure 7.1.

		Pitfall	Amidation at COOH	Me at COOH	Underprop	Overprop	Reversed overprop
Sample prep.		Method	A, F	A, A.n	B, C, A.r, A.t, H 20x, H 5x	D, E, F, G, H, H 42x, H20x	H 42x boil, H42x ha
Data acquisition	IMPLICATIONS	No LC retention	-	-	X	-	-
		MS precursor signal ↓	X	X	X	X	-
		Chimeric MSMS spectra ↑	X	-	-	-	-
		MSMS redundancy ↑	X	X	x	X	-
Data analysis	ID	No ID of Ac at S, T or Y	-	-	-	-	X
		ID of semi-ArgC peptides	-	-	X	-	-
		Combinatorial explosion ↑	X	X	X	X	-
		Chimeric spectra hinder ID	X	-	-	-	-
	Quant	Co-elution hinders quantification	X	-	-	-	-
		No relative inter run quantification	X	X	X	X	-

Figure 7.1 Overview of the implications for data acquisition and data analysis, related to several pitfalls that can emerge during propionylation of histones

First, amidation of carboxyl groups can occur when using ammonium hydroxide and/or ABC as a buffer system during the propionylation or digestion procedure. Nevertheless, not all methods using these products showed amidation, such as is the case for e.g. method B and C. This mainly depends on the order in which the propionylation reagent and buffer system are added. If propionic anhydride can react with the carboxyl groups in the histone sample and form a "mixed anhydride", prior to the addition of ammonium hydroxide, the latter will react with this "mixed anhydride" resulting in an amidated carboxyl group (Figure 6.2). This leads to the distribution of one former MS precursor over several amidated forms thereof. Consequently, these less intense peptide forms can co-elute, leading to chimeric spectra and take up valuable time during MSMS acquisition. Moreover, the subsequent identification and quantification will be hindered as well.

Another unwanted side reaction that can occur at carboxyl groups is methylation. This reaction occurs when methanol reacts with the "mixed anhydride" peptide form in stead of ammonium hydroxide. However, this side effect is of minor relevance as the reaction with ammonium hydroxide is favored.

Next, inefficient propionylation was observed in several methods. The main reason for this is the presence of other primary amines (coming from ammonium hydroxide or ABC), that are competing for propionylation with histones. A second cause can be lowering the molar excess of propionic anhydride from 420x to 20x or lower, which also results in underpropionylated peptides. Just as for amidation, underpropionylation can result in the original precursor being distributed over several peptide forms, although no chimeric spectra are expected since there is no co-elution of identical m/z ratio's. Nevertheless, an additional problem is the inadequate LC retention of non-propionylated tryptic peptides. The identification of *in vivo* PTMs will be hampered by the additional amount of variable "*in vitro* induced" PTMs and by impaired enzyme specificity. Finally, the inconsistent derivatization efficiency in between samples makes relative quantification unreliable.

Further, aspecific propionylation at hydroxyl groups of S, T and Y residues will result in the same issues as underpropionylation, except for the decrease in LC retention and increase in semi-ArgC specificity. Breaking this ester bond by boiling or hydroxylamine mediated acyl removal will undo all of these complications, as depicted in Figure 7.1. Nevertheless, this is at the cost of identifying one specific type of biological PTM at these residues: acetylation. As the chemical bond of the latter is the same as for the chemically induced propionylgroup, both will most probably be removed simultaneously. Still, we believe this method is arguably a very good candidate since the acetylation of S and T is a sproradic event compared to other PTMs which can now be more easily identified and quantified.

In conclusion, chemical derivatization of histones prior to bottom-up MS is not as straight forward as it was considered to be. Many pitfalls can interfere with data acquisition as well as data analysis, which were already quite demanding to start with, given the large amount of PTMs present on these histones. After examining various approaches, we put forward a new propionylation strategy whereby overpropionylation on S, T and Y residues is reversed by the addition of hydroxylamine, resulting in only correctly propionylated peptides. This protocol will facilitate both identification and quantification of all biological histone PTMs – apart from acetylation at S and T.

"Prediction is very difficult, especially about the future"

— Niels Bohr

CHAPTER 8:

BROADER INTERNATIONAL CONTEXT, RELEVANCE AND FUTURE PERSPECTIVES

8. BROADER INTERNATIONAL CONTEXT, RELEVANCE AND FUTURE PERSPECTIVES

"We now have the possibility of achieving all we ever hoped for from medicine". This statement of the UK science minister in June 2000 seemed to mark a new era, where the origin of each disease could be uncovered. The trigger for this remarkable announcement was the report that the Human Genome Project public consortium assembled a working draft of the sequence of the human genome - the genetic blueprint of a human being. Over a decade and thousands of human genomes later, this statement has proven at best premature, since reality turned out to be somewhat more complicated.

Epigenetics might be able to fill up at least some of the remaining gaps. The term epigenetics, which was coined by Waddington in 1942, was derived from the Greek word "epigenesis". Waddington presented his metaphorical "epigenetic landscape" to exemplify the concepts of developmental biology. Nowadays, it includes all heritable changes in gene expression (active versus inactive genes) that do not involve changes to the underlying DNA sequence; a change in phenotype without a change in genotype. Biologists offer the following metaphor as an explanation: if the genome is the hardware, then the epigenome is the software. Or as Joseph Ecker, a Salk Institute biologist puts it: "I can load Windows, if I want, on my Mac. You're going to have the same chip in there, the same genome, but different software. And the outcome is a different cell type."

Epigenetics is one of the fastest growing areas in science these days and is now a key issue in studies concerning development and disease. The amount of publications per year concerning this topic have increased from about 300 around the turn of the century, to over 7000 for the year 2015. While many epigenetic scientists have devoted a great part of their research to DNA methylation, in the meantime also many other different types of epigenetic processes were identified. Histone PTMs are one of them. The combinatorial nature of these histone modifications reveals a "histone code" that considerably extends the information potential of the genetic code. Just like for epigenetics in general, the amount of publications on histone PTMs in specific has massively increased from a humble 72 in the year 2000 to over 1700 the past year. Throughout this dissertation, we tried to color one small area of the biological picture related to the histone code (histone clipping in hESC) and also made some suggestions to strengthen the technical procedures that provide the framework for studying biology (chemical derivatization prior to bottom-up MS analysis).

First, we zoomed in on histone H3 proteolysis as a new type of PTM in hESC. We found this event to be dependent on the conditions hESC were cultured in (feeder-free or MEF feeder layer). This is in accordance with the findings in mESC, where the cleavage pattern altered during embryonic body formation [184]. Only very recently the group of Jensen published a paper where they state histones

H2B and H3 undergo proteolytic processing in primary human hepatocytes and the hepatocellular carcinoma cell line HepG2/C3A when grown in spheroid (3D) culture, but not in a flat (2D) culture [226]. These examples underline the relevance of our findings and the need for further research, as currently a full understanding of this process is lacking. In general, the concept of histone proteolysis has had a long history of disregard ever since the first reports occured during the 70s and 80s. The revival started around 2010 with the first evidence of epigenetic potential of histone clipping in yeast and mouse [184,191][253]. However, up to now this phenomenom of histone proteolysis is still not being picked up by the broader scientific community. Moreover, the papers that do appear, often do not indicate whether degradation or clipping was studied. Given the importance of histone modifications in epigenetic regulation, we proposed to divide histone proteolysis into histone degradation and the epigenetically connoted histone clipping. As such, readers can hopefully find their way more easily through the confusing amalgam of recent reports in which findings of often little biological coherence are continuously being cross-referenced.

Prudence in drawing immediate biological conclusion is also called for during the study of other PTMs than histone proteolysis. Mass spectrometry currently is the only technique holding the promise of simultaneously annotating and quantifying enough different hPTM combinations to considerably mine the histone code to a depth that is required for interpreting the biology underlying this histone code. The second part of this thesis however surfaces several pitfalls in the standard methodology applied in MS-based histone study, as we came across different types of unexpected PTMs that were induced during chemical derivatization of histones for subsequent bottom-up MS analysis. Our results have implications on data analysis for all MS based histone research and more specifically on quantitation: current findings are only correct when using relative abundances as a metric to find changes between different samples. The relevance of this notion was stressed in a commentary article on our findings, written by the group of Garcia [224]. Because of these and other inconveniences - such as loss of MSMS acquisition time on *in vitro* generated precursors – we developed a new strategy for derivatizing these histones. This propionylation procedure is now efficient as well as specific, with the only drawback being that acetylations on S, T or Y residues can no longer be identified. We hope this derivatization protocol will be implemented by the growing amount of research groups using bottom-up MS when investigating histones.

When looking at future perspectives, we would like to distinguish between the goals of our own research group in particular and the more broad-shouldered ambitions of epigenetic research in general. At present, our research group is focusing on further deciphering the histone code in differentiating hESC. **(i)** On the one hand, experiments are carried out to identify the enzyme mediating the previously

disclosed H3 clipping in hESC. Therefore, the nuclear extract of clipping hESC is fractionated and the different fractions are tested for their H3 clipping ability. Subsequently all fractions will be analyzed by MS in order to detect an enzyme unique for the clipping fractions. **(ii)** On the other hand, we will continue investigating histone PTMs through MS. In order to mine the histone epigenome of differentiating hESC, we will apply two new DIA strategies, HDMS^E (Waters) and SWATH (Sciex). In a first step, we will focus on data analysis of histone PTMs in terms of identification and quantification. Some very concrete challeges still remain for data analysis platforms to be addressed concerning MS identification of histone PTMs. First, mass spectra need to be preprocessed, which can be difficult for middle down and top down anlysis due to the limitations in resolution. Moreover, chimeric spectra originating from isobaricly modified peptides make interpration at the MSMS level more complex. The next challenge is to increase the identification rate at a certain FDR. The wide variety of PTMs can lead to a combinatorial explosion and in increasing risk of false positive identifications. How to fast and accurately process them is the key for accurate identification. Finally, correct localization of a PTM remains difficult, as typically not all fragment ion are available. Subsequently, the DIA analysis platform will be used to start developing an "epigenetic toxicology screening test" for monitoring the changes in histone PTMs upon treatment with different pharmacological compounds. Toxic agents could act at different levels, either directly on enzymes controlling epigenetic changes or indirectly to activate or block signalling pathways causing long-term alterations in chromatin structure that may or may not be heritable. The extensive list of drugs withdrawn from the market because of unwanted side effect calls for some consideration [254]. A broad descriptive analysis made clear that among 121 prescription drugs, 42.1% were withdrawn from the European markets alone between 1960 and 1999 [256]. The current preclinical and clinical studies often do not succeed in tracing all possible side effect, not even taking into account the effects on pregnant women or children that are hardly ever investigated. Therefore it is important broaden our knowledge on the epigenetic changes involved when cells or organisms are exposed to chemicals. With the increasing introduction of DNA-methylation patterns in toxicology screenings, it might be that histone modification analysis makes its entry in the near future [255]. With the inclusion of hESC, the guidelines formulated by "*Toxicity Testing in the 21st Century: A Vision and a Strategy*" are taken into account as well [257]. By means of developing a DIA MS based screening test our research groups aims at further extending the knowledge on toxicology induced changes in histone PTMs. The final goal is for research groups and drug developers to be able to anticipate possible side effect of new drug targets.

These objectives partly overlay with the goals set by the NIH Roadmap Epigenomics Mapping Consortium, launched in 2008. This latter project set out to generate a public resource of human epigenomic data from 111 different types of human cells and tissues: stem cells, mature cells from a

variety of different tissues from healthy people and from patients with diseases such as cancer, neurodegenerative conditions and autoimmune disease. In February 2015, the journal Nature published a suite of papers that serve as a joint update of the consortium's findings and the ongoing progress. In general, these findings collectively provide insights into three fundamental aspects of epigenetics: how the epigenome affects gene expression, the epigenetic changes during stem cell differentiation, and the interplay between epigenetics and disease. Over the next 7 to 10 years, the consortium aims to expand on this work by deciphering 1000 human epigenomes. So far, it is still hard to predict where this epigenome roadmap will lead us, but joint efforts might bring the statement of the UK Science minster in 2000 to live after all.

CHAPTER 9:

SUMMARY

9. SUMMARY

All cells in the human body contain the same DNA. In contrast, they do not all fulfill the same function. Hence, it is clear that DNA is not our only destiny. In the 1940's, Waddington brought up the term "epigenotype" for the first time [32] and ever since it has been used to connect genotype and environment to phenotype and disease. This thesis focusses on histone-related epigenetics from two different perspectives. On the one hand, we investigated the biology of histone cleavage as a potential epigenetic mechanisms during hESC differentiation. On the other, the technical challenges involved in studying posttranslational histone modifications through bottom-up mass spectrometry (MS) are addressed, whereupon suggestions for better sample preparation are made.

Chapter 1 provides the biological framework as a basis for enabling the interpretation of the following chapters of this thesis. Thereby we begin by discussing the origin of hESC, available culture methods, differentiation possibilities and characterization of this specific cell type. Next, we put forward two different types of stem cells and consider the transition from bench to bedside from an ethical point of view. This section on stem cells is followed by a brief definition of the term epigenetics and the categorization into the different biological processes involved in the epigenetic landscape and changes thereof. Because of the ample technical optimization for MS analysis on histone PTMs described in this thesis, **Chapter 2** offers a technical background. Some basic principles of proteomics in general and of MS specifically are discussed, followed by a section dedicated to the different possible MS strategies for studying histone PTMs.

Being different from most studies regarding histone PTMs, the first part of the research (**Chapter 4**) does not describe the specific introduction or removal of a single functional group on the amino acid chain, but focuses on a phenomenon called histone proteolysis, in the setting of early stem cell differentiation. This type of PTM has received only little attention in the field even though its occurrence may lead to drastic epigenetic changes. If an amino- or carboxy-terminal histone tail gets clipped, all accompanying PTMs, both transcription-activating and -inhibiting ones, are erased simultaneously. Chapter 4 outlines in more detail that also in hESC the aminoterminal tail of histone H3 can be cleaved. We investigated the effect of the culture condition on this events and report that clipping of histone H3 can be found both in feeder-free as well as MEF feeder layer culturing condition, albeit in a dissimilar pattern. Whilst the H3 clipping was a continuous process for feeder-free cultured stem cells, it appeared to be more of a fluctuating event if these hESC were cultured on a MEF feeder layer during differentiation. The reason for this discrepancy is not clear, but it might be explained by the difference in composition, in hESC colony architecture, in cell signaling, and in the biomechanics of the cellular microenvironment. All our subsequent attempts to disclose this clipping event as an *in vitro* artefact did confirm our assumption

that this event takes place prior to extraction, in the nucleus itself. Next, we also report that there is no direct link between H3 clipping and the expression of Oct4 - a hallmark pluripotency factor – or the cell cycle. By the use of different complementary techniques, we then found three different cleavage sites: A21 (western blotting), R26 (Edman degradation) and residue 31 (MS). Finally, by means of a protease inhibitor panel, we were able to identify the enzyme responsible for the clipping event as a serine protease. To date, the lack of evidence for transcriptional implications of H3 clipping during hESC differentiation calls for prudence in categorizing H3 proteolysis as an epigenetic event. Therefore, as described in the **first part of Chapter 7,** a proposal is given to divide histone proteolysis into histone degradation on the one hand and the epigenetically connoted histone clipping on the other.

Next to the biological complexity involved in histone epigenetics, there are also various technical challenges. MS stands at the heart of proteome research and is also the method of choice for the large-scale PTM-analysis. However, the most commonly used strategy - bottom-up MS analysis - requires derivatization of lysine residues prior to the trypsin digest, in order to obtain sufficient hydrophobic peptides of an appropriate length. Generally, the primary amines are propionylated by use of propionic anhydride. Unfortunately, the technical variation introduced by this additional step during sample preparation might hinder drawing the correct biological conclusions. In this context, we focused on the issues concerning both propionylation efficiency as well as specificity that need to be overcome.

In **Chapter 5** we examined different propionylation strategies found in literature and encountered several side reactions hindering the specificity or efficiency of the derivatization process. The main pitfalls are amidation and methylation of carboxy residues, aspecific overpropionylation on S, T and Y and inefficient propionylation of the primary amines on K-residues as well as peptide N-termini. In order to address each of these issues, we made several adjustments to these methods and developed some new propionylation strategies in **Chapter 6**. Unfortunately, many adjustments made seemed to exchange one problem for another, rendering most methods unfit for future MS analysis. Nevertheless, one specific strategy does meet the criteria of a specific and efficient propionylation strategy. In this method, overpropionylation on all monitored peptides was reversed by hydroxylamine-mediated acyl removal, resulting in correctly propionylated peptides. The average conversion rate of this approach was 95%. The only drawback when using this method is that the acetylation at S and/or T residues can no longer be identified or quantified, as this ester bond is broken as well. This appears to be the small price we need to pay to be able to identify and quantify all other PTMs that in a more reliable manner. The implications on data acquisition and data analysis of all methods investigated are considered in the **second part of Chapter 7.**

To sum up, this dissertation has unraveled some layers of histone proteomics both from a biological and a technical point of view. In the first section, we were able to characterize histone H3 clipping in hESC. In the second part, we disclosed various pitfalls during chemical derivatization of histones prior to bottom-up MS analysis and put forward an optimized method for future analysis. However, to get to the core of true histone epigenetics, many layers still need to be peeled off. On the bright side, each contribution can be considered an additional pen stroke to design an epigenome roadmap, as directed by the US National Institutes of Health. Let's hope this roadmap will be able to guide us through the epigenetic jungle in the near future, and provide crucial information to connect genetic variation and disease.

HOOFDSTUK 10:

SAMENVATTING

10. SAMENVATTING

Alle cellen in een menselijk lichaam beschikken over hetzelfde DNA, niettegenstaande ze niet allemaal dezelfde functie vervullen. Klaarblijkelijk beslist het DNA dus niet als enige over de eindbestemming van een cel. Rond 1940 opperde Waddington voor het eerst de term "epigenotype" [32] en van toen af aan gebruikt men dit begrip om de link te maken vanuit genotype en omgeving naar fenotype en ziekte. Deze thesis focust op histon gerelateerde epigenetica vanuit twee verschillende invalshoeken. Enerzijds stellen we een onderzoek in naar de biologische context van histon klieving als een mogelijk epigenetisch proces tijdens stemceldifferentiatie. Anderzijds leggen we ons toe op de technische uitdagingen die gepaard gaan met het bestuderen van posttranslationele histon modificaties (PTMs) aan de hand van bottom-up massaspectrometrie (MS), waarbij we enkele suggesties maken voor een betere staalvoorbereiding.

Hoofdstuk 1 omvat de biologische omkadering die dient als basis om de daaropvolgende hoofdstukken te interpeteren. Daarbij starten we met het bespreken van de herkomst van humane embryonale stamcellen (hESC), de beschikbare cultuursystemen, de differentiatiemogelijkheden en de karakterisatie van dit speciefieke celtype. Vervolgens brengen we twee verschillende types stamcellen ter sprake en beschouwen we de overgang vanuit de ondezoekslaboratoria naar de patient vanuit een ethisch oogpunt. Dit onderdeel rond stamcellen wordt gevolgd door een korte definitie van de term epigenetica en de onderverdeling daarvan in verscheidene biologische processen die betrekking hebben op dit epigentische landschap en de veranderingen ervan. Gezien de uitvoerige beschrijving van de technische optimalisatie voor MS analyse in dit werk, biedt **hoofdstuk 2** de nodige technische achtergrond. Enkele basisprincipes van eiwitonderzoek in het algemeen en MS in het bijzonder worden hierin toegelicht, alsook de verschillende benaderingen om histon PTMs te bestuderen aan de hand van MS.

In tegenstelling tot de meeste studies naar histon PTMs, legt het eerste deel van dit onderzoek (**hoofdstuk 4**) zich niet toe op de specifieke aan- dan wel afwezigheid van een bepaalde functionele groep op de keten van aminozuren. Het focust daarentegen op een fenomeen tijdens vroege stamceldifferentiatie dat histon proteolyse wordt genoemd. Dit type PTM kreeg tot op heden nog maar weinig aandacht, ondanks de grote epigentische veranderingen die het met zich mee kan brengen. Als de histonstaart amino- of carboxyterminaal ingekort wordt dan zullen namelijk gelijktijdig alle bijhorende PTMs op deze staarten verwijderd worden. Hoofdstuk 4 beschrijft in detail hoe zo een N-terminale klieving van histon H3 ook in hESC kan plaatsvinden. We onderzochten het effect van cultuurcondities op dit process en rapporteren dat histon H3 kllieving zich zowel mét als zonder MEF feeder layer manifesteert, zij het volgens een ander patroon. Hoewel deze klieving doorlopend plaatsgrijpt in een stamcelcultuur zonder feeder laag, blijkt de aanwezigheid van MEFs als feeder laag

ervoor te zorgen dat dit proces een golvend patroon volgt tijdens differentiatie. De oorzaak voor deze discrepantie is nog niet gekend, maar een mogelijke verklaring is het verschil in compositie, architectuur, cel signalisatie en biomechanica in het micromilieu rondom de cellen. Alle voorafgaandelijke pogingen om aan te tonen dat deze klieving zich *in vitro* voordoet, bevestigden ons vermoeden dat dit verschijnsel wel degelijk in de celkern zelf plaatsvindt, voorafgaandelijk aan extractie. Vervolgens rapporteren we tevens dat noch de expressie van Oct4 – een pluripotentiemerker – noch de celcyclus in verband kon worden gebracht met deze H3 klieving. Aan de hand van verscheidene complementaire technieken konden we drie verschillende klievingsplaatsen aanduiden: A21 (western blot), R26 (Edman degradatie) en residue 31 (MS). Ten slotte hebben we met behulp van een verzameling aan protease inhibitoren kunnen aantonen dat het enzyme verantwoordelijk voor deze klieving tot de serine proteases behoort. Echter, tot op heden dwingt het gebrek aan bewijzen betreffende het belang van H3 klieving op transcriptie-niveau ons tot voorzichtigheid wanneer we H3 proteolyse als een epigenetisch concept beschouwen. Daarom doen we in het **eerste deel van hoofdstuk 7** een voorstel tot het onderverdelen van histon proteolyse in "histon afbraak" enerzijds en een meer epigenetisch getinte "histon klieving" anderzijds.

Naast de biologische complexiteit van histon epigentica, brengt het ook verschillende technische uitdagingen met zich mee. MS vormt de basis van eiwitonderzoek en is tevens eerste keuze bij de grootschalige analyse van PTMs. Echter, om de meest gebruikte strategie – bottom-up MS analyse- te hanteren is het noodzakelijk de lysines (K) chemisch te derivatiseren voorafgaandelijk aan incubatie met trypsine, opdat er voldoende hydrofobe peptiden van geschikte lengte zouden gegenereerd worden. Meestal worden de primaire amines gederivatiseerd door middel van propionzuuranhydride. Helaas wordt er heel wat technische variatie geïntroduceerd tijdens deze stap in de staalvoorbereiding, wat het trekken van de juiste biologische conclusies kan hinderen. Binnen deze context, legden we ons toe op de problemen gerelateerd aan zowel de efficiëntie als specificiteit van deze propionylatiereactie.

In **hoofdstuk 5** bestudeerden we verscheidene propionylatie methoden die terug te vinden zijn in de literatuur en stootten daarbij op ettelijke nevenreacties die zowel de efficiëntie als de specificiteit van deze derivatisaties kunnen verstoren. De voornaamste struikelblokken zijn amidatie en methylatie van carboxylgroepen, aspecifieke overpropionylatie op serine (S) threonine (T) en tyrosine (Y), en inefficiënte propionylatie van de primaire amines op K en de N-termini van peptiden. Om aan deze problemen tegemoet te komen werden verschillende aanpassingen doorgevoerd en eveneens enkele nieuwe propionylatie methodes ontwikkeld, zoals beschreven in **hoofdstuk 6.** Jammer genoeg zorgden heel wat aanpassingen voor het uitwisselen van het ene probleem met een ander, waardoor de methode ongeschikt bleef voor verder gebruik. Desalniettemin voldoet één methode aan de vooropgestelde

criteria betreffende efficiëntie en specificiteit. Hierbij wordt de overpropionylatie op alle bestudeerde peptiden teniet gedaan onder invloed van hydroxylamine gestuurde verwijdering van de acyl groep, waardoor enkel nog correct gepropionyleerde peptides overbijven. De gemiddelde derivatisatiegraad voor deze methode bedraagt meer dan 95%. Het enige minpunt bij het aanwenden van deze strategie is dat acetylatie op S en/of T residuen niet langer geïndentifceerd of gekwantificeerd kan worden, gezien de ester binding tussen beide eveneens verbroken zal worden. Dit blijkt echter het kleine offer dat we moeten maken opdat alle andere PTMs op een meer betrouwbare wijze geïdentificeerd en gekwantificeerd kunnen worden. De consequenties ten aanzien van data acquisitie en data analyse worden voor alle methoden verder uitgediept in het **tweede deel van hoofdstuk 7.**

Kortom, deze thesis heeft binnen het histoneiwit-onderzoek verschillende knopen ontward, zowel vanuit een biologisch als technisch oogpunt. In het eerste deel omschrijven we histon H3 klieving in hESC. Doorheen het tweede deel leggen we verscheidene struikelblokken bloot tijdens de chemische derivatisatie van histonen voorafgaand aan bottom-up MS analyse en dragen daarbij een nieuwe geoptimaliseerde methode aan voor toekomstige analyses. Desalniettemin, als men de histoncode volledig wil ontcijferen zullen nog heel wat bijkomende knopen ontward moeten worden. Het goede nieuws is dat elke bijdrage beschouwd kan worden als een bijkomende pennentrek om de "epigenetische wegenkaart" te ontwerpen, onder leiding van "the US National Institutes of health". De hoop leeft dat deze kaart ons in de nabije toekomst in staat zal stellen doorheen de epigentische jungle te navigeren en bijgevolg cruciale informatie kan onthullen om genetische variatie en ziekte aan mekaar te verbinden.

CHAPTER 11:

ADDENDUM

11. ADDENDUM

11.1. Supporting information Chapter 5

Table 11.1 Propionylation methods used in Chapter 5

	Method						
	A		**B**		**C**		**D**
Substrate (vacuum dried bovine histones)	10 µg		10 µg		10 µg		10 µg
Propionylation reagent	MeOH:prop.anh (3:1)		MeOH:prop.anh (3:1)		MeOH:prop.anh (3:1)		100 mM NHS-propionate solution:ACN (1:1)
Propionylation reaction	+ 20 µL prop. reagent + 15 µL NH_4OH		+ 30 µL 50 mM ABC + 10 µL NH_4OH + 10 µL prop. Reagent adjust pH with NH_4OH (8-10)		+ 30 µL 50 mM ABC + 10 µL NH_4OH + 10 µL prop. Reagent adjust pH (8-10)		+ 50 mM TEAB + prop. Reagent at 10 mM Incubate * + prop. Reagent at 10 mM Incubate * + NH_2OH, 10 min at 30°C adjust pH (3)
Incubation	30 min at RT		15 min at 37°C		15 min at 51°C		1 h at 30°C *
	A.1	A.2	B.1	B.2	C.1	C.2	1x pre digest 1x post digest
Frequency	1x pre digest 1x post digest	2x pre digest 2x post digest	1x pre digest 1x post digest	2x pre digest 2x post digest	1x pre digest 1x post digest	2x pre digest 2x post digest	

MeOH = methanol; prop. anh = propionic anhydride; prop. reagent = propionylation reagent; ABC = ammoniumbicarbonate; ACN = acetonitrile; TEAB = triethylammoniumbicarbonate

11.2. Supporting information Chapter 6

Table 11.2 Propionylation methods used in Chapter 6

	Method			
	A.r	**A.t**	**A.n**	**B.s**
Substrate (vacuum dried bovine histones)	10 µg	10 µg	10 µg	10 µg
Propionylation reagent	MeOH:prop.anh (3:1)	MeOH:prop.anh:NH_4OH (3:1:3)	MeOH:prop.anh (3:1)	Prop. anh
Propionylation reaction	+ 15 µL NH_4OH + 20 µL prop. reagent	+ 35 µL prop. reagent	+ 20 µL prop. reagent + 20 µL H_2O	+ 10 µL 25 mM ABC + 10 µL MeOH + 10 µL prop. reagent + 20 µL NH_4OH (pH 8)
Incubation	30 min at RT	30 min at RT	30 min at RT	30 min at RT
Frequency	2x pre digest 2x post digest	2x pre digest 2x post digest	2x pre digest 2x post digest	2x pre digest 2x post digest

MeOH = methanol; prop. anh = propionic anhydride; prop. reagent = propionylation reagent; TEAB = triethylammonium bicarbonate

	Method			
	E	**F**	**G**	**H**
Substrate (vacuum dried bovine histones)	10 µg	10 µg	10 µg	20 µg
Propionylation reagent	Prop. anh	Prop. anh	Prop. anh	IPA:prop.anh (3:1)
Propionylation reaction	+ 50 µL propionic acid + 5µL prop. anh	+ 15 µL 0.1 M ethanolamine + 5 µL prop.anh	+ 15 µL 0.1 M triethanolamine + 5 µL prop.anh	20 µL 1.0 M TEAB + 20 µL prop. reagent Incubate * + 20 µL H_2O, 30 min at 37°C
Incubation	6h at RT	30 min at RT	30 min at RT	30 min at RT *
Frequency	2x pre digest 2x post digest	2x pre digest 2x post digest	2x pre digest 2x post digest	2x pre digest 2x post digest

IPA = isopropylalcohol; prop. anh = propionic anhydride; prop. reagent = propionylation reagent; TEAB = triethylammoniumbicarbonate

	Method		
	H 42x	**H 20x**	**H 5x**
Substrate (vacuum dried bovine histones)	20 µg	20 µg	20 µg
Propionylation reagent	IPA:prop.anh (79:1)	IPA:prop.anh (170:1)	IPA:prop.anh (680:1)
Propionylation reaction	20 µL 1.0 M TEAB + 20 µL prop. reagent Incubate * + 20 µL H_2O, 30 min at 37°C	20 µL 1.0 M TEAB + 20 µL prop. reagent Incubate * + 20 µL H_2O, 30 min at 37°C	20 µL 1.0 M TEAB + 20 µL prop. reagent Incubate * + 20 µL H_2O, 30 min at 37°C
Incubation	30 min at RT *	30 min at RT *	30 min at RT *
Frequency	1x pre digest 1x post digest	1x pre digest 1x post digest	1x pre digest 1x post digest

IPA = isopropylalcohol; prop. anh = propionic anhydride; prop. reagent = propionylation reagent; TEAB = triethylammonium bicarbonate

	Method for reversing overpropionylation	
	Boiling mediated	**Hydroxylamine mediated**
Reaction	Vacuum dry sample + 50 µL 50 mM ABC adjust pH with NH_4OH (8-10)	Vacuum dry sample + 50 µL 0.5 M NH_2OH + 15 µL NH_4OH at pH 12 Incubate * Adjust pH with formic acid (3)
Incubation	1 h at 99°C	20 min at RT *

ABC = ammoniumbicarbonate

REFERENCES

[1] Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., et al., Embryonic stem cell lines derived from human blastocysts. *Science (80-.).* 1998, 282, 1145–1147.

[2] Braam, S.R., Zeinstra, L., Litjens, S., Ward-van Oostwaard, D., et al., Recombinant vitronectin is a functionally defined substrate that supports human embryonic stem cell self-renewal via alphavbeta5 integrin. *Stem Cells* 2008, 26, 2257–2265.

[3] Villa-Diaz, L.G., Pacut, C., Slawny, N.A., Ding, J., et al., Analysis of the factors that limit the ability of feeder cells to maintain the undifferentiated state of human embryonic stem cells. *Stem Cells Dev.* 2009, 18, 641–651.

[4] Winslow, T., Duckwall, C., Techniques for Generating Embryonic Stem Cell Cultures. *2011* n.d.

[5] Murry, C.E., Keller, G., Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. *Cell* 2008, 132, 661–80.

[6] Williams, L.A., Davis-Dusenbery, B.N., Eggan, K.C., SnapShot: directed differentiation of pluripotent stem cells. *Cell* 2012, 149, 1174–1174.e1.

[7] Blanpain, C., Horsley, V., Fuchs, E., Epithelial stem cells: turning over new leaves. *Cell* 2007, 128, 445–58.

[8] Arrin, Bongso, L.E.H., *Stem Cells: From Bench to Bedside*, World Scientific, 2011.

[9] Draper, J.S., Pigott, C., Thomson, J.A., Andrews, P.W., Surface antigens of human embryonic stem cells: changes upon differentiation in culture. *J Anat* 2002, 200, 249–258.

[10] Sperger, J.M., Chen, X., Draper, J.S., Antosiewicz, J.E., et al., Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. *Proc. Natl. Acad. Sci. U. S. A.* 2003, 100, 13350–5.

[11] Van Hoof, D., Mummery, C.L., Heck, A.J., Krijgsveld, J., Embryonic stem cell proteomics. *Expert Rev Proteomics* 2006, 3, 427–437.

[12] Takahashi, K., Yamanaka, S., Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 2006, 126, 663–676.

[13] Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., et al., Induced pluripotent stem cell lines derived from human somatic cells. *Science (80-.).* 2007, 318, 1917–1920.

[14] González, F., Boué, S., Izpisúa Belmonte, J.C., Methods for making induced pluripotent stem cells: reprogramming à la carte. *Nat. Rev. Genet.* 2011, 12, 231–42.

[15] Stadtfeld, M., Hochedlinger, K., Induced pluripotency: history, mechanisms, and applications. *Genes Dev.* 2010, 24, 2239–2263.

[16] Stadtfeld, M., Nagaya, M., Utikal, J., Weir, G., Hochedlinger, K., Induced pluripotent stem cells generated without viral integration. *Science* 2008, 322, 945–9.

[17] Diecke, S., Jung, S.M., Lee, J., Ju, J.H., Recent technological updates and clinical applications of induced pluripotent stem cells. *Korean J. Intern. Med.* 2014, 29, 547–57.

[18] Zhou, H., Wu, S., Joo, J.Y., Zhu, S., et al., Generation of induced pluripotent stem cells using

recombinant proteins. *Cell Stem Cell* 2009, 4, 381–4.

[19] Jia, F., Wilson, K.D., Sun, N., Gupta, D.M., et al., A nonviral minicircle vector for deriving human iPS cells. *Nat. Methods* 2010, 7, 197–9.

[20] Warren, L., Manos, P.D., Ahfeldt, T., Loh, Y.-H., et al., Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell* 2010, 7, 618–30.

[21] Nichols, J., Smith, A., Naive and primed pluripotent states. *Cell Stem Cell* 2009, 4, 487–492.

[22] Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., et al., New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* 2007, 448, 196–9.

[23] Brons, I.G.M., Smithers, L.E., Trotter, M.W.B., Rugg-Gunn, P., et al., Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* 2007, 448, 191–5.

[24] O'Leary, T., Heindryckx, B., Lierman, S., van Bruggen, D., et al., Tracking the progression of the human inner cell mass during embryonic stem cell derivation. *Nat. Biotechnol.* 2012, 30, 278–282.

[25] O'Leary, T., Heindryckx, B., Lierman, S., Van der Jeught, M., et al., Derivation of human embryonic stem cells using a post-inner cell mass intermediate. *Nat. Protoc.* 2013, 8, 254–64.

[26] Van der Jeught, M., O'Leary, T., Duggal, G., De Sutter, P., et al., The post-inner cell mass intermediate: implications for stem cell biology and assisted reproductive technology. *Hum. Reprod. Update* 2015, 21, 616–26.

[27] Zhang, B., Krawetz, R., Rancourt, D.E., Would the real human embryonic stem cell please stand up? *BioEssays* 2013, 35, 632–638.

[28] Devolder, K., *The Ethics of Embryonic Stem Cell Research*, OUP Oxford, 2015.

[29] Lo, B., Parham, L., Ethical issues in stem cell research. *Endocr. Rev.* 2009, 30, 204–13.

[30] King, N.M., Perrin, J., Ethical issues in stem cell research and therapy. *Stem Cell Res. Ther.* 2014, 5, 85.

[31] Liu, W., Deng, Y., Liu, Y., Gong, W., Deng, W., Stem cell models for drug discovery and toxicology studies. *J. Biochem. Mol. Toxicol.* 2013, 27, 17–27.

[32] Waddington, C.H., The epigenotype. 1942. *Int. J. Epidemiol.* 2012, 41, 10–13.

[33] Russo, V.E.A., Martienssen, R.A., Riggs, A.D., Epigenetic mechanisms of gene regulation. 1996.

[34] Goldberg, A.D., Allis, C.D., Bernstein, E., Epigenetics: a landscape takes shape. *Cell* 2007, 128, 635–8.

[35] Ladewig, J., Koch, P., Brüstle, O., Leveling Waddington: the emergence of direct programming and the loss of cell fate hierarchies. *Nat. Rev. Mol. Cell Biol.* 2013, 14, 225–236.

[36] Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., Richmond, T.J., Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* 1997, 389, 251–260.

[37] Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., et al., Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* 1998, 393, 386–

389.

[38] Popp, C., Dean, W., Feng, S., Cokus, S.J., et al., Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* 2010, 463, 1101–1105.

[39] He, Y.F., Li, B.Z., Li, Z., Liu, P., et al., Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science (80-.).* 2011, 333, 1303–1307.

[40] Bhutani, N., Burns, D.M., Blau, H.M., DNA demethylation dynamics. *Cell* 2011, 146, 866–872.

[41] Dr Adele Murrell, DNA Methylation and Genome Stability. *DNA Methylation Genome Stab.* n.d.

[42] The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* 2004, 306, 636–40.

[43] Kaikkonen, M.U., Lam, M.T.Y., Glass, C.K., Non-coding RNAs as regulators of gene expression and epigenetics. *Cardiovasc. Res.* 2011, 90, 430–40.

[44] Ernst, C., Morton, C.C., Identification and function of long non-coding RNA. *Front. Cell. Neurosci.* 2013, 7, 168.

[45] Krol, J., Loedige, I., Filipowicz, W., The widespread regulation of microRNA biogenesis, function and decay. *Nat. Rev. Genet.* 2010, 11, 597–610.

[46] Collins, L., Schönfeld, B., Chen, X.S., in:, Tollefsbol T (Ed.), *Handb. Epigenetics new Mol. Med. Genet.*, Academic Press, 2011, pp. 49–61.

[47] Liebers, R., Rassoulzadegan, M., Lyko, F., Epigenetic regulation by heritable RNA. *PLoS Genet.* 2014, 10, e1004296.

[48] Ostermeier, G.C., Dix, D.J., Miller, D., Khatri, P., Krawetz, S.A., Spermatozoal RNA profiles of normal fertile men. *Lancet (London, England)* 2002, 360, 772–7.

[49] Krawetz, S.A., Paternal contribution: new insights and future challenges. *Nat. Rev. Genet.* 2005, 6, 633–42.

[50] Walser, C.B., Lipshitz, H.D., Transcript clearance during the maternal-to-zygotic transition. *Curr. Opin. Genet. Dev.* 2011, 21, 431–43.

[51] Ostermeier, G.C., Miller, D., Huntriss, J.D., Diamond, M.P., Krawetz, S.A., Reproductive biology: delivering spermatozoan RNA to the oocyte. *Nature* 2004, 429, 154.

[52] Grandjean, V., Badro, D.A., Kiani, J., RNA: a possible contributor to the "missing heritability". *Basic Clin. Androl.* 2013, 23, 9.

[53] Narlikar, G.J., Sundaramoorthy, R., Owen-Hughes, T., Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell* 2013, 154, 490–503.

[54] Lessard, J., Crabtree, G.R., Chromatin Regulatory Mechanisms in Pluripotency. *Annu Rev Cell Dev Biol* 2010, 26, 503–532.

[55] Tollervey, J.R., Lunyak, V. V, Epigenetics: judge, jury and executioner of stem cell fate. *Epigenetics* 2012, 7, 823–840.

[56] Martin, C., Zhang, Y., Mechanisms of epigenetic inheritance. *Curr. Opin. Cell Biol.* 2007, 19, 266–72.

[57] Kumari, S., Swaminathan, A., Chatterjee, S., Senapati, P., et al., Chromatin organization, epigenetics and differentiation: an evolutionary perspective. *Subcell. Biochem.* 2013, 61, 3–35.

[58] Gan, Q., Yoshida, T., McDonald, O.G., Owens, G.K., Concise review: epigenetic mechanisms contribute to pluripotency and cell lineage determination of embryonic stem cells. *Stem Cells* 2007, 25, 2–9.

[59] Tonna, S., El-Osta, A., Cooper, M.E., Tikellis, C., Metabolic memory and diabetic nephropathy: potential role for epigenetic mechanisms. *Nat. Rev. Nephrol.* 2010, 6, 332–41.

[60] Huang, H., Sabari, B.R., Garcia, B.A., Allis, C.D., Zhao, Y., SnapShot: Histone Modifications. *Cell* 2014, 159, 458–458.e1.

[61] Tan, M., Luo, H., Lee, S., Jin, F., et al., Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* 2011, 146, 1016–1028.

[62] Wedemeyer, W.J., Welker, E., Scheraga, H.A., Proline cis-trans isomerization and protein folding. *Biochemistry* 2002, 41, 14637–44.

[63] Jenuwein, T., Allis, C.D., Translating the histone code. *Science (80-.).* 2001, 293, 1074–1080.

[64] Bassett, A., Cooper, S., Wu, C., Travers, A., The folding and unfolding of eukaryotic chromatin. *Curr Opin Genet Dev* 2009, 19, 159–165.

[65] Ruthenburg, A.J., Li, H., Patel, D.J., Allis, C.D., Multivalent engagement of chromatin modifications by linked binding modules. *Nat. Rev. Mol. Cell Biol.* 2007, 8, 983–994.

[66] Huang, H., Lin, S., Garcia, B.A., Zhao, Y., Quantitative Proteomic Analysis of Histone Modifications. *Chem. Rev.* 2015.

[67] Khare, S.P., Habib, F., Sharma, R., Gadewal, N., et al., HIstome--a relational knowledgebase of human histone proteins and histone modifying enzymes. *Nucleic Acids Res.* 2012, 40, D337–42.

[68] Kouzarides, T., Chromatin modifications and their function. *Cell* 2007, 128, 693–705.

[69] Kühn, S., Hofmeyr, J.-H.S., Is the "Histone Code" an Organic Code? *Biosemiotics* 2014, 7, 203–222.

[70] Herceg, Z., Murr, R., in:, Tollefsbol T (Ed.), *Handb. Epigenetics new Mol. Med. Genet.*, Academic Press, 2011, pp. 25–45.

[71] Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., et al., Active genes are tri-methylated at K4 of histone H3. *Nature* 2002, 419, 407–11.

[72] Briggs, S.D., Bryk, M., Strahl, B.D., Cheung, W.L., et al., Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae. *Genes Dev.* 2001, 15, 3286–95.

[73] Bernstein, B.E., Humphrey, E.L., Erlich, R.L., Schneider, R., et al., Methylation of histone H3 Lys 4 in coding regions of active genes. *Proc. Natl. Acad. Sci. U. S. A.* 2002, 99, 8695–700.

[74] Swigut, T., Wysocka, J., H3K27 demethylases, at long last. *Cell* 2007, 131, 29–32.

[75] Zhang, Y., Transcriptional regulation by histone ubiquitination and deubiquitination. *Genes Dev* 2003, 17, 2733–2740.

[76] Iñiguez-Lluhí, J.A., For a healthy histone code, a little SUMO in the tail keeps the acetyl away. *ACS Chem. Biol.* 2006, 1, 204–6.

[77] Johnson, E.S., Protein modification by SUMO. *Annu. Rev. Biochem.* 2004, 73, 355–82.

[78] Kamakaka, R.T., Biggins, S., Histone variants: deviants? *Genes Dev* 2005, 19, 295–310.

[79] Jufvas, A., Stralfors, P., Vener, A. V, Histone variants and their post-translational modifications in primary human fat cells. *PLoS One* 2011, 6, e15960.

[80] Meshorer, E., Misteli, T., Chromatin in pluripotent embryonic stem cells and differentiation. *Nat Rev Mol Cell Biol* 2006, 7, 540–546.

[81] Ausio, J., Histone variants--the structure behind the function. *Br. Funct Genomic Proteomic* 2006, 5, 228–243.

[82] Zalensky, A.O., Siino, J.S., Gineitis, A.A., Zalenskaya, I.A., et al., Human testis/sperm-specific histone H2B (hTSH2B). Molecular cloning and characterization. *J. Biol. Chem.* 2002, 277, 43474–80.

[83] Churikov, D., Siino, J., Svetlova, M., Zhang, K., et al., Novel human testis-specific histone H2B encoded by the interrupted gene on the X chromosome. *Genomics* 2004, 84, 745–56.

[84] Ahmad, K., Henikoff, S., The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Mol Cell* 2002, 9, 1191–1200.

[85] Ng, R.K., Gurdon, J.B., Epigenetic memory of an active gene state depends on histone H3.3 incorporation into chromatin in the absence of transcription. *Nat Cell Biol* 2008, 10, 102–109.

[86] Smith, L.M., Kelleher, N.L., Consortium for Top Down, P., Proteoform: a single term describing protein complexity. *Nat. Methods* 2013, 10, 186–187.

[87] Twyman, R.M., *Principles of Proteomics*, Garland Science, 2013.

[88] Angel, T.E., Aryal, U.K., Hengel, S.M., Baker, E.S., et al., Mass spectrometry-based proteomics: existing capabilities and future directions. *Chem. Soc. Rev.* 2012, 41, 3912–28.

[89] Marx, V., Targeted proteomics. *Nat. Methods* 2012, 10, 19–22.

[90] Ahmad, Y., Lamond, A.I., A perspective on proteomics in cell biology. *Trends Cell Biol.* 2014, 24, 257–64.

[91] SciCurve n.d.

[92] Domon, B., Aebersold, R., Options and considerations when selecting a quantitative proteomics strategy. *Nat. Biotechnol.* 2010, 28, 710–721.

[93] Yamashita, M., Fenn, J.B., Electrospray ion source. Another variation on the free-jet theme. *J. Phys. Chem.* 1984, 88, 4451–4459.

[94] Shevchenko, A., Wilm, M., Vorm, O., Mann, M., Mass Spectrometric Sequencing of Proteins from Silver-Stained Polyacrylamide Gels. *Anal. Chem.* 1996, 68, 850–858.

[95] Banerjee, S., Mazumdar, S., Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. *Int. J. Anal. Chem.* 2012, 2012, 282574.

[96] Andrews, G.L., Simons, B.L., Young, J.B., Hawkridge, A.M., Muddiman, D.C., Performance characteristics of a new hybrid quadrupole time-of-flight tandem mass spectrometer (TripleTOF 5600). *Anal. Chem.* 2011, 83, 5442–6.

[97] Helm, D., Vissers, J.P.C., Hughes, C.J., Hahne, H., et al., Ion mobility tandem mass spectrometry enhances performance of bottom-up proteomics. *Mol. Cell. Proteomics* 2014, 13, 3709–15.

[98] Waters, No Title n.d.

[99] Syka, J.E.P., Coon, J.J., Schroeder, M.J., Shabanowitz, J., Hunt, D.F., Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* 2004, 101, 9528–33.

[100] Biemann, K., Contributions of mass spectrometry to peptide and protein structure. *Biomed. Environ. Mass Spectrom.* 1988, 16, 99–111.

[101] Roepstorff, P., Fohlman, J., Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.* 1984, 11, 601.

[102] Hart-Smith, G., Blanksby, S., Mass analysis. *Fac. Sci. Med. Heal. - Pap.* 2012, 5 – 30.

[103] Colaert, N., Digging deeper in the data, novel approaches and analysis methods in quantitative peptide-centric proteomics 2011.

[104] Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., et al., Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* 2002, 1, 376–86.

[105] Bendall, S.C., Hughes, C., Stewart, M.H., Doble, B., et al., Prevention of amino acid conversion in SILAC experiments with embryonic stem cells. *Mol. Cell. Proteomics* 2008, 7, 1587–97.

[106] Bicho, C.C., de Lima Alves, F., Chen, Z.A., Rappsilber, J., Sawin, K.E., A genetic engineering solution to the "arginine conversion problem" in stable isotope labeling by amino acids in cell culture (SILAC). *Mol. Cell. Proteomics* 2010, 9, 1567–77.

[107] Geiger, T., Cox, J., Ostasiewicz, P., Wisniewski, J.R., Mann, M., Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nat. Methods* 2010, 7, 383–385.

[108] Krijgsveld, J., Ketting, R.F., Mahmoudi, T., Johansen, J., et al., Metabolic labeling of C. elegans and D. melanogaster for quantitative proteomics. *Nat. Biotechnol.* 2003, 21, 927–31.

[109] Doherty, M.K., Whitehead, C., McCormack, H., Gaskell, S.J., Beynon, R.J., Proteome dynamics in complex organisms: using stable isotopes to monitor individual protein turnover rates. *Proteomics* 2005, 5, 522–33.

[110] Krüger, M., Moser, M., Ussar, S., Thievessen, I., et al., SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. *Cell* 2008, 134, 353–64.

[111] Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., et al., Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* 1999, 17, 994–9.

[112] Boersema, P.J., Raijmakers, R., Lemeer, S., Mohammed, S., Heck, A.J.R., Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat. Protoc.* 2009, 4, 484–94.

[113] Zhang, R., Sioma, C.S., Wang, S., Regnier, F.E., Fractionation of isotopically labeled peptides in quantitative proteomics. *Anal. Chem.* 2001, 73, 5142–9.

[114] Stewart, I.I., Thomson, T., Figeys, D., 18O labeling: a tool for proteomics. *Rapid Commun. Mass Spectrom.* 2001, 15, 2456–65.

[115] Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., et al., Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* 2004, 3, 1154–1169.

[116] Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., et al., Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Anal. Chem.* 2003, 75, 1895–1904.

[117] Karp, N.A., Huber, W., Sadowski, P.G., Charles, P.D., et al., Addressing accuracy and precision issues in iTRAQ quantitation. *Mol. Cell. Proteomics* 2010, 9, 1885–1897.

[118] Bantscheff, M., Lemeer, S., Savitski, M.M., Kuster, B., Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal. Bioanal. Chem.* 2012, 404, 939–965.

[119] Evans, C., Noirel, J., Ow, S.Y., Salim, M., et al., An insight into iTRAQ: where do we stand now? *Anal. Bioanal. Chem.* 2012.

[120] Lundgren, D.H., Hwang, S.-I., Wu, L., Han, D.K., Role of spectral counting in quantitative proteomics. *Expert Rev. Proteomics* 2010, 7, 39–53.

[121] Gillet, L.C., Navarro, P., Tate, S., Rost, H., et al., Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol. Cell. Proteomics* 2012, 11.

[122] Silva, J.C., Gorenstein, M. V, Li, G.-Z., Vissers, J.P.C., Geromanos, S.J., Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol. Cell. Proteomics* 2006, 5, 144–56.

[123] Geromanos, S.J., Vissers, J.P.C., Silva, J.C., Dorschel, C.A., et al., The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. *Proteomics* 2009, 9, 1683–1695.

[124] Gerber, S.A., Rush, J., Stemman, O., Kirschner, M.W., Gygi, S.P., Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. U. S. A.* 2003, 100, 6940–6945.

[125] Britton, L.-M.P., Gonzales-Cope, M., Zee, B.M., Garcia, B.A., Breaking the histone code with quantitative mass spectrometry. *Expert Rev. Proteomics* 2011, 8, 631–643.

[126] Soldi, M., Cuomo, A., Bremang, M., Bonaldi, T., Mass spectrometry-based proteomics for the analysis of chromatin structure and dynamics. *Int J Mol Sci* 2013, 14, 5402–5431.

[127] Garcia, B.A., Shabanowitz, J., Hunt, D.F., Characterization of histones and their post-translational modifications by mass spectrometry. *Curr. Opin. Chem. Biol.* 2007, 11, 66–73.

[128] Garcia, B.A., Mollah, S., Ueberheide, B.M., Busby, S.A., et al., Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nat. Protoc.* 2007, 2, 933–938.

[129] Liao, R., Wu, H., Deng, H., Yu, Y., et al., Specific and efficient N-propionylation of histones with propionic acid N-hydroxysuccinimide ester for histone marks characterization by LC-MS. *Anal. Chem.* 2013, 85, 2253–2259.

[130] Smith, C.M., Haimberger, Z.W., Johnson, C.O., Wolf, A.J., et al., Heritable chromatin structure: mapping "memory" in histones H3 and H4. *Proc. Natl. Acad. Sci. U. S. A.* 2002, 99 Suppl 4, 16454–61.

[131] Smith, C.M., Gafken, P.R., Zhang, Z., Gottschling, D.E., et al., Mass spectrometric quantification of acetylation at specific lysines within the amino-terminal tail of histone H4. *Anal. Biochem.* 2003, 316, 23–33.

[132] Moradian, A., Kalli, A., Sweredoski, M.J., Hess, S., The top-down, middle-down, and bottom-up mass spectrometry approaches for characterization of histone variants and their post-translational modifications. *Proteomics* 2013.

[133] Siuti, N., Kelleher, N.L., Decoding protein modifications using top-down mass spectrometry. *Nat. Methods* 2007, 4, 817–821.

[134] Karch, K.R., Denizio, J.E., Black, B.E., Garcia, B.A., Identification and interrogation of combinatorial histone modifications. *Front. Genet.* 2013, 4, 264.

[135] Sidoli, S., Cheng, L., Jensen, O.N., Proteomics in chromatin biology and epigenetics: Elucidation of post-translational modifications of histone proteins by mass spectrometry. *J Proteomics* 2012, 75, 3419–3433.

[136] DiMaggio, P.A., Young, N.L., Baliban, R.C., Garcia, B.A., Floudas, C.A., A Mixed Integer Linear Optimization Framework for the Identification and Quantification of Targeted Post-translational Modifications of Highly Modified Proteins Using Multiplexed Electron Transfer Dissociation Tandem Mass Spectrometry. *Mol. Cell. Proteomics* 2009, 8, 2527–2543.

[137] Frank, A.M., Pesavento, J.J., Mizzen, C.A., Kelleher, N.L., Pevzner, P.A., Interpreting Top-Down Mass Spectra Using Spectral Alignment. *Anal. Chem.* 2008, 80, 2499–2505.

[138] Guan, S., Burlingame, A.L., Data processing algorithms for analysis of high resolution MSMS spectra of peptides with complex patterns of posttranslational modifications. *Mol. Cell. Proteomics* 2010, 9, 804–10.

[139] McLafferty, F.W., Breuker, K., Jin, M., Han, X., et al., Top-down MS, a powerful complement to the high capabilities of proteolysis proteomics. *FEBS J.* 2007, 274, 6256–68.

[140] Steen, H., Mann, M., The ABC's (and XYZ's) of peptide sequencing. *Nat. Rev. Mol. Cell Biol.* 2004, 5, 699–711.

[141] Sadygov, R.G., Cociorva, D., Yates, J.R., Large-scale database searching using tandem mass spectra: looking up the answer in the back of the book. *Nat. Methods* 2004, 1, 195–202.

[142] Baliban, R.C., DiMaggio, P.A., Plazas-Mayorca, M.D., Young, N.L., et al., A novel approach for untargeted post-translational modification identification using integer linear optimization and tandem mass spectrometry. *Mol. Cell. Proteomics* 2010, 9, 764–79.

[143] Hansen, B.T., Davey, S.W., Ham, A.-J.L., Liebler, D.C., P-Mod: an algorithm and software to map modifications to peptide sequences using tandem MS data. *J. Proteome Res.* 4, 358–68.

[144] Tsur, D., Tanner, S., Zandi, E., Bafna, V., Pevzner, P.A., Identification of post-translational modifications by blind search of mass spectra. *Nat. Biotechnol.* 2005, 23, 1562–7.

[145] Tanner, S., Shu, H., Frank, A., Wang, L.-C., et al., InsPecT: identification of posttranslationally modified peptides from tandem mass spectra. *Anal. Chem.* 2005, 77, 4626–39.

[146] Chen, Y., Chen, W., Cobb, M.H., Zhao, Y., PTMap--a sequence alignment software for unrestricted, accurate, and full-spectrum identification of post-translational modification sites. *Proc. Natl. Acad. Sci. U. S. A.* 2009, 106, 761–6.

[147] Fu, Y., Qian, X., Transferred Subgroup False Discovery Rate for Rare Post-translational Modifications Detected by Mass Spectrometry. *Mol. Cell. Proteomics* 2014, 13, 1359–68.

[148] Huang, X., Huang, L., Peng, H., Guru, A., et al., ISPTM: an iterative search algorithm for systematic identification of post-translational modifications from complex proteome mixtures. *J. Proteome Res.* 2013, 12, 3831–42.

[149] Yuan, Z.-F., Lin, S., Molden, R.C., Garcia, B.A., Evaluation of proteomic search engines for the analysis of histone modifications. *J. Proteome Res.* 2014, 13, 4470–8.

[150] Marshall, A.G., Hendrickson, C.L., High-resolution mass spectrometers. *Annu. Rev. Anal. Chem. (Palo Alto. Calif).* 2008, 1, 579–599.

[151] Beausoleil, S.A., Villén, J., Gerber, S.A., Rush, J., Gygi, S.P., A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat. Biotechnol.* 2006, 24, 1285–92.

[152] Cox, J., Mann, M., MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 2008, 26, 1367–72.

[153] Savitski, M.M., Lemeer, S., Boesche, M., Lang, M., et al., Confident phosphorylation site localization using the Mascot Delta Score. *Mol. Cell. Proteomics* 2011, 10, M110.003830.

[154] Ruttenberg, B.E., Pisitkun, T., Knepper, M.A., Hoffert, J.D., PhosphoScore: an open-source phosphorylation site assignment tool for MSn data. *J. Proteome Res.* 2008, 7, 3054–9.

[155] Albuquerque, C.P., Smolka, M.B., Payne, S.H., Bafna, V., et al., A multidimensional chromatography technology for in-depth phosphoproteome analysis. *Mol. Cell. Proteomics* 2008, 7, 1389–96.

[156] Bailey, C.M., Sweet, S.M.M., Cunningham, D.L., Zeller, M., et al., SLoMo: automated site localization of modifications from ETD/ECD mass spectra. *J. Proteome Res.* 2009, 8, 1965–71.

[157] Taus, T., Köcher, T., Pichler, P., Paschke, C., et al., Universal and confident phosphorylation site localization using phosphoRS. *J. Proteome Res.* 2011, 10, 5354–62.

[158] Phanstiel, D.H., Brumbaugh, J., Wenger, C.D., Tian, S., et al., Proteomic and phosphoproteomic comparison of human ES and iPS cells. *Nat. Methods* 2011, 8, 821–7.

[159] Baker, P.R., Trinidad, J.C., Chalkley, R.J., Modification site localization scoring integrated into a search engine. *Mol. Cell. Proteomics* 2011, 10, M111.008078.

[160] Vaudel, M., Breiter, D., Beck, F., Rahnenführer, J., et al., D-score: a search engine independent MD-score. *Proteomics* 2013, 13, 1036–41.

[161] Dai, L., Peng, C., Montellier, E., Lu, Z., et al., Lysine 2-hydroxyisobutyrylation is a widely distributed active histone mark. *Nat. Chem. Biol.* 2014, 10, 365–70.

[162] Park, J., Chen, Y., Tishkoff, D.X., Peng, C., et al., SIRT5-mediated lysine desuccinylation impacts diverse metabolic pathways. *Mol. Cell* 2013, 50, 919–30.

[163] Bonenfant, D., Towbin, H., Coulot, M., Schindler, P., et al., Analysis of dynamic changes in post-translational modifications of human histones during cell cycle by mass spectrometry. *Mol. Cell. Proteomics* 2007, 6, 1917–1932.

[164] Jung, H.R., Pasini, D., Helin, K., Jensen, O.N., Quantitative mass spectrometry of histones H3.2 and H3.3 in Suz12-deficient mouse embryonic stem cells reveals distinct, dynamic post-translational modifications at Lys-27 and Lys-36. *Mol. Cell. Proteomics* 2010, 9, 838–50.

[165] Cuomo, A., Moretti, S., Minucci, S., Bonaldi, T., SILAC-based proteomic analysis to dissect the "histone modification signature" of human breast cancer cells. *Amino Acids* 2011, 41, 387–99.

[166] Guan, X., Rastogi, N., Parthun, M.R., Freitas, M.A., Discovery of histone modification crosstalk networks by stable isotope labeling of amino acids in cell culture mass spectrometry (SILAC MS). *Mol. Cell. Proteomics* 2013, 12, 2048–59.

[167] Hammond, S.L., Byrum, S.D., Namjoshi, S., Graves, H.K., et al., Mitotic phosphorylation of histone H3 threonine 80. *Cell Cycle* 2014, 13, 440–52.

[168] Zhao, Y., Fang, X., Wang, Y., Zhang, J., et al., Comprehensive analysis for histone acetylation of human colon cancer cells treated with a novel HDAC inhibitor. *Curr. Pharm. Des.* 2014, 20, 1866–73.

[169] Zhang, K., Schrag, M., Crofton, A., Trivedi, R., et al., Targeted proteomics for quantification of histone acetylation in Alzheimer's disease. *Proteomics* 2012, 12, 1261–8.

[170] Weinert, B.T., Iesmantavicius, V., Moustafa, T., Schölz, C., et al., Acetylation dynamics and stoichiometry in Saccharomyces cerevisiae. *Mol. Syst. Biol.* 2014, 10, 716.

[171] Glibert, P., Vossaert, L., Van Steendam, K., Lambrecht, S., et al., Quantitative Proteomics to Characterize Specific Histone H2A Proteolysis in Chronic Lymphocytic Leukemia and the Myeloid THP-1 Cell Line. *Int. J. Mol. Sci.* 2014, 15, 9407–9421.

[172] Yuan, X., Gu, X., Crabb, J.S., Yue, X., et al., Quantitative proteomics: comparison of the macular Bruch membrane/choroid complex from age-related macular degeneration and normal eyes. *Mol. Cell. Proteomics* 2010, 9, 1031–46.

[173] Montes de Oca, R., Shoemaker, C.J., Gucek, M., Cole, R.N., Wilson, K.L., Barrier-to-autointegration factor proteome reveals chromatin-regulatory partners. *PLoS One* 2009, 4, e7050.

[174] Salim, K., Kehoe, L., Minkoff, M.S., Bilsland, J.G., et al., Identification of differentiating neural progenitor cell markers using shotgun isobaric tagging mass spectrometry. *Stem Cells Dev.* 2006, 15, 461–70.

[175] Pesavento, J.J., Yang, H., Kelleher, N.L., Mizzen, C.A., Certain and progressive methylation of histone H4 at lysine 20 during the cell cycle. *Mol. Cell. Biol.* 2008, 28, 468–86.

[176] Scharf, A.N.D., Meier, K., Seitz, V., Kremmer, E., et al., Monomethylation of lysine 20 on histone H4 facilitates chromatin maturation. *Mol. Cell. Biol.* 2009, 29, 57–67.

[177] Zee, B.M., Levin, R.S., Dimaggio, P.A., Garcia, B.A., Global turnover of histone post-translational modifications and variants in human cells. *Epigenetics Chromatin* 2010, 3, 22.

[178] Ong, S.-E., Mittler, G., Mann, M., Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. *Nat. Methods* 2004, 1, 119–26.

[179] Lin, S., Garcia, B.A., Examining histone posttranslational modification patterns by high-resolution mass spectrometry. *Methods Enzym.* 2012, 512, 3–28.

[180] Distler, U., Kuharev, J., Navarro, P., Levin, Y., et al., Drift time-specific collision energies enable deep-coverage data-independent acquisition proteomics. *Nat. Methods* 2014, 11, 167–70.

[181] Lin, S., Wein, S., Gonzales-Cope, M., Otte, G.L., et al., Stable-isotope-labeled histone peptide library for histone post-translational modification and variant quantification by mass spectrometry. *Mol. Cell. Proteomics* 2014, 13, 2450–66.

[182] Bernstein, E., Hake, S.B., The nucleosome: a little variation goes a long way. *Biochem Cell Biol* 2006, 84, 505–517.

[183] Skene, P.J., Henikoff, S., Histone variants in pluripotency and disease. *Development* 2013, 140, 2513–2524.

[184] Duncan, E.M., Muratore-Schroeder, T.L., Cook, R.G., Garcia, B.A., et al., Cathepsin L proteolytically processes histone H3 during mouse embryonic stem cell differentiation. *Cell* 2008, 135, 284–294.

[185] Duncan, E.M., Allis, C.D., Errors in erasure: links between histone lysine methylation removal and disease. *Prog. Drug Res.* 2011, 67, 69–90.

[186] Adams-Cioaba, M.A., Krupa, J.C., Xu, C., Mort, J.S., Min, J., Structural basis for the recognition and cleavage of histone H3 by cathepsin L. *Nat Commun* 2011, 2, 197.

[187] Tesar, M., Marquardt, O., Foot-and-mouth disease virus protease 3C inhibits cellular transcription and mediates cleavage of histone H3. *Virology* 1990, 174, 364–374.

[188] Falk, M.M., Grigera, P.R., Bergmann, I.E., Zibert, A., et al., Foot-and-mouth disease virus protease 3C induces specific proteolytic cleavage of host cell histone H3. *J Virol* 1990, 64, 748–756.

[189] Gonzalo, S., Epigenetic alterations in aging. *J Appl Physiol* 2010, 109, 586–597.

[190] Mahendra, G., Kanungo, M.S., Age-related and steroid induced changes in the histones of the quail liver. *Arch Gerontol Geriatr* 2000, 30, 109–114.

[191] Santos-Rosa, H., Kirmizis, A., Nelson, C., Bartke, T., et al., Histone H3 tail clipping regulates gene expression. *Nat. Struct. Mol. Biol.* 2009, 16, 17–22.

[192] Allis, C.D., Bowen, J.K., Abraham, G.N., Glover, C. V, Gorovsky, M.A., Proteolytic processing of histone H3 in chromatin: a physiologically regulated event in Tetrahymena micronuclei. *Cell* 1980, 20, 55–64.

[193] Allis, C.D., Wiggins, J.C., Proteolytic processing of micronuclear H3 and histone phosphorylation during conjugation in Tetrahymena thermophila. *Exp Cell Res* 1984, 153, 287–298.

[194] Mandal, P., Azad, G.K., Tomar, R.S., Identification of a novel histone H3 specific protease activity in nuclei of chicken liver. *Biochem. Biophys. Res. Commun.* 2012, 421, 261–7.

[195] Mandal, P., Verma, N., Chauhan, S., Tomar, R.S., Unexpected histone H3 tail-clipping activity of glutamate dehydrogenase. *J. Biol. Chem.* 2013, 288, 18743–18757.

[196] Zwaka, T.P., Thomson, J.A., Homologous recombination in human embryonic stem cells. *Nat. Biotechnol.* 2003, 21, 319–321.

[197] Zhang, C., Liu, Y., Andrews, P.C., Quantification of histone modifications using ^{15}N metabolic labeling. *Methods* 2013, 61, 236–43.

[198] Glibert, P., Van Steendam, K., Dhaenens, M., Deforce, D., iTRAQ as a method for optimization: enhancing peptide recovery after gel fractionation. *Proteomics* 2014, 14, 680–684.

[199] Vossaert, L., O Leary, T., Van Neste, C., Heindryckx, B., et al., Reference loci for RT-qPCR analysis of differentiating human embryonic stem cells. *BMC Mol Biol* 2013, 14, 21.

[200] Scheerlinck, E., Van Steendam, K., Vandewoestyne, M., Lepez, T., et al., Detailed method description for noninvasive monitoring of differentiation status of human embryonic stem cells. *Anal. Biochem.* 2014, 461, 60–66.

[201] Johns, E.W., Studies on histones. 7. Preparative methods for histone fractions from calf thymus. *Biochem. J.* 1964, 92, 55–59.

[202] Phillips, D.M., Johns, E.W., A study of the proteinase content and the chromatography of thymus histones. *Biochem. J.* 1959, 72, 538–544.

[203] Gafni, O., Weinberger, L., Mansour, A.A., Manor, Y.S., et al., Derivation of novel human ground state naive pluripotent stem cells. *Nature* 2013, 504, 282–286.

[204] Azad, G.K., Tomar, R.S., Proteolytic clipping of histone tails: the emerging role of histone proteases in regulation of various biological processes. *Mol Biol Rep* 2014, 41, 2717–2730.

[205] Xue, Y., Vashisht, A.A., Tan, Y., Su, T., Wohlschlegel, J.A., PRB1 Is Required for Clipping of the Histone H3 N Terminal Tail in Saccharomyces cerevisiae. *PLoS One* 2014, 9, e90496.

[206] Strahl, B.D., Allis, C.D., The language of covalent histone modifications. *Nature* 2000, 403, 41–45.

[207] Xu, G., Shin, S.B., Jaffrey, S.R., Global profiling of protease cleavage sites by chemoselective labeling of protein N-termini. *Proc. Natl. Acad. Sci. U. S. A.* 2009, 106, 19310–19315.

[208] Zhang, C., Liu, Y., Andrews, P.C., Quantification of histone modifications using N metabolic labeling. *Methods* 2013.

[209] Dawson, M.A., Kouzarides, T., Cancer epigenetics: from mechanism to therapy. *Cell* 2012, 150, 12–27.

[210] Baylin, S.B., Jones, P.A., A decade of exploring the cancer epigenome - biological and translational implications. *Nat. Rev. Cancer* 2011, 11, 726–34.

[211] Chi, P., Allis, C.D., Wang, G.G., Covalent histone modifications--miswritten, misinterpreted and mis-erased in human cancers. *Nat. Rev. Cancer* 2010, 10, 457–69.

[212] Feinberg, A.P., Epigenetic stochasticity, nuclear structure and cancer: the implications for medicine. *J. Intern. Med.* 2014, 276, 5–11.

[213] Helin, K., Dhanak, D., Chromatin proteins and modifications as drug targets. *Nature* 2013, 502, 480–488.

[214] Fraenkel-Conrat, H., OLCOTT, H.S., Esterification of proteins with alcohols of low molecular weight. *J. Biol. Chem.* 1945, 161, 259–268.

[215] *Applications of Mass Spectrometry in Life Safety*, Springer Science & Business Media, 2008.

[216] McDonald, L., Robertson, D.H.L., Hurst, J.L., Beynon, R.J., Positional proteomics: selective recovery and analysis of N-terminal proteolytic peptides. *Nat. Methods* 2005, 2, 955–7.

[217] Davidson, G.R., Armstrong, S.D., Beynon, R.J., Positional proteomics at the N-terminus as a means of proteome simplification. *Methods Mol. Biol.* 2011, 753, 229–42.

[218] Fojo, A.T., Reuben, P.M., Whitney, P.L., Awad, W.M., Effect of glycerol on protein acetylation by acetic anhydride. *Arch. Biochem. Biophys.* 1985, 240, 43–50.

[219] Zhang, Q., Qian, W.-J., Knyushko, T. V, Clauss, T.R.W., et al., A method for selective enrichment and analysis of nitrotyrosine-containing peptides in complex proteome samples. *J. Proteome Res.* 2007, 6, 2257–68.

[220] Abello, N., Kerstjens, H.A.M., Postma, D.S., Bischoff, R., Selective acylation of primary amines in peptides and proteins. *J. Proteome Res.* 2007, 6, 4770–6.

[221] Wiktorowicz, J.E., English, R.D., Wu, Z., Kurosky, A., Model studies on iTRAQ modification of peptides: sequence-dependent reaction specificity. *J. Proteome Res.* 2012, 11, 1512–20.

[222] Pflieger, D., Junger, M.A., Muller, M., Rinner, O., et al., Quantitative proteomic analysis of protein complexes: concurrent identification of interactors and their state of phosphorylation. *Mol. Cell. Proteomics* 2008, 7, 326–346.

[223] Münchbach, M., Quadroni, M., Miotto, G., James, P., Quantitation and facilitated de novo sequencing of proteins by isotopic N-terminal labeling of peptides with a fragmentation-directing moiety. *Anal. Chem.* 2000, 72, 4047–57.

[224] Sidoli, S., Garcia, B.A., Properly reading the histone code by MS-based proteomics. *Proteomics* 2015, 15, 2901–2902.

[225] Orford, K.W., Scadden, D.T., Deconstructing stem cell self-renewal: genetic insights into cell-cycle regulation. *Nat Rev Genet* 2008, 9, 115–128.

[226] Tvardovskiy, A., Wrzesinski, K., Sidoli, S., Fey, S.J., et al., Top-down and middle-down protein analysis reveals that intact and clipped human histones differ in post-translational modification patterns. *Mol. Cell. Proteomics* 2015.

[227] Burdick, J.A., Vunjak-Novakovic, G., Engineered microenvironments for controlled stem cell differentiation. *Tissue Eng. Part A* 2009, 15, 205–19.

[228] Nava, M.M., Raimondi, M.T., Pietrabissa, R., Controlling self-renewal and differentiation of stem cells via mechanical cues. *J. Biomed. Biotechnol.* 2012, 2012, 797410.

[229] Qian, M.-X., Pang, Y., Liu, C.H., Haratake, K., et al., Acetylation-mediated proteasomal degradation of core histones during DNA repair and spermatogenesis. *Cell* 2013, 153, 1012–1024.

[230] Lin, R., Cook, R.G., Allis, C.D., Proteolytic removal of core histone amino termini and dephosphorylation of histone H1 correlate with the formation of condensed chromatin and transcriptional silencing during Tetrahymena macronuclear development. *Genes Dev* 1991, 5, 1601–1610.

[231] Kaul, R., Hoang, A., Yau, P., Bradbury, E.M., Wenman, W.M., The chlamydial EUO gene encodes a histone H1-specific protease. *J. Bacteriol.* 1997, 179, 5928–34.

[232] Allis, C.D., Glover, C. V, Gorovsky, M.A., Micronuclei of Tetrahymena contain two types of histone H3. *Proc. Natl. Acad. Sci. U. S. A.* 1979, 76, 4857–61.

[233] Capozzo, A.V.E., Burke, D.J., Fox, J.W., Bergmann, I.E., et al., Expression of foot and mouth disease virus non-structural polypeptide 3ABC induces histone H3 cleavage in BHK21 cells. *Virus Res.* 2002, 90, 91–9.

[234] Grigera, P.R., Tisminetzky, S.G., Histone H3 modification in BHK cells infected with foot-and-mouth disease virus. *Virology* 1984, 136, 10–9.

[235] Zhou, P., Wu, E., Alam, H.B., Li, Y., Histone cleavage as a mechanism for epigenetic regulation: current insights and perspectives. *Curr. Mol. Med.* 2014, 14, 1164–72.

[236] Mishra, R.N., Kanungo, M.S., Alterations in histones of the liver and oviduct of Japanese quail during aging. *Mol Biol Rep* 1994, 20, 15–18.

[237] Haugen, M.H., Johansen, H.T., Pettersen, S.J., Solberg, R., et al., Nuclear legumain activity in colorectal cancer. *PLoS One* 2013, 8, e52980.

[238] Khalkhali-Ellis, Z., Goossens, W., Margaryan, N. V, Hendrix, M.J.C., Cleavage of histone 3 by cathepsin d in the involuting mammary gland. *PLoS One* 2014, 9, e103230.

[239] Lane, A.A., Ley, T.J., Neutrophil elastase cleaves PML-RARalpha and is important for the development of acute promyelocytic leukemia in mice. *Cell* 2003, 115, 305–18.

[240] Goulet, B., Baruch, A., Moon, N.S., Poirier, M., et al., A cathepsin L isoform that is devoid of a signal peptide localizes to the nucleus in S phase and processes the CDP/Cux transcription factor. *Mol Cell* 2004, 14, 207–219.

[241] Goulet, B., Sansregret, L., Leduy, L., Bogyo, M., et al., Increased expression and activity of nuclear cathepsin L in cancer cells suggests a novel mechanism of cell transformation. *Mol. Cancer Res.* 2007, 5, 899–907.

[242] Goulet, B., Markovic, Y., Leduy, L., Nepveu, A., Proteolytic processing of cut homeobox 1 by neutrophil elastase in the MV4;11 myeloid leukemia cell line. *Mol. Cancer Res.* 2008, 6, 644–53.

[243] Uy, G.L., Lane, A.A., Welch, J.S., Grieselhuber, N.R., et al., A protease-resistant PML-RAR{alpha} has increased leukemogenic potential in a murine model of acute promyelocytic leukemia. *Blood* 2010, 116, 3604–10.

[244] Mayer, K., Vreemann, A., Qu, H., Brix, K., Release of endo-lysosomal cathepsins B, D, and L from IEC6 cells in a cell culture model mimicking intestinal manipulation. *Biol. Chem.* 390, 471–80.

[245] Butler, G.S., Overall, C.M., Proteomic identification of multitasking proteins in unexpected locations complicates drug targeting. *Nat. Rev. Drug Discov.* 2009, 8, 935–948.

[246] Kirschke, H., Langner, J., Wiederanders, B., Ansorge, S., Bohley, P., Cathepsin L. A new proteinase from rat-liver lysosomes. *Eur J Biochem* 1977, 74, 293–301.

[247] Benes, P., Vetvicka, V., Fusek, M., Cathepsin D--many functions of one aspartic protease. *Crit. Rev. Oncol. Hematol.* 2008, 68, 12–28.

[248] Zubenko, G.S., Jones, E.W., Protein degradation, meiosis and sporulation in proteinase-deficient mutants of Saccharomyces cerevisiae. *Genetics* 1981, 97, 45–64.

[249] Khairalla, A.S., Omer, S.A., Mahdavi, J., Aslam, A., et al., Nuclear trafficking, histone cleavage and

induction of apoptosis by the meningococcal App and MspA autotransporters. *Cell. Microbiol.* 2015, 17, 1008–20.

[250] Arnandis, T., Ferrer-Vicens, I., Torres, L., García, C., et al., Differential functions of calpain 1 during epithelial cell death and adipocyte differentiation in mammary gland involution. *Biochem. J.* 2014, 459, 355–368.

[251] Kim, H.G., Kim, K.C., Roh, T.Y., Park, J., et al., Gene silencing in HIV-1 latency by polycomb repressive group. *Virol J* 2011, 8, 179.

[252] Vossaert, L., Meert, P., Scheerlinck, E., Glibert, P., et al., Identification of histone H3 clipping activity in human embryonic stem cells. *Stem Cell Res.* 2014, 13, 123–134.

[253] Duncan, E.M., PhD Thesis: Regulated Histone H3 Proteolysis During Mouse Embryonic Stem Cell Differentiation. Rockefeller University, 2009.

[254] Geneesmiddelen die uit de handel zijn genomen omwille van ongewenste effecten: enkele reflecties n.d.

[255] Priestley, C.C., Anderton, M., Doherty, A.T., Duffy, P., et al., Epigenetics – relevance to drug safety science. *Toxicol. Res. (Camb).* 2012, 1, 23.

[256] Fung, M., Thornton, A., Mybeck, K., Wu, J.H. -h., et al., Evaluation of the Characteristics of Safety Withdrawal of Prescription Drugs from Worldwide Pharmaceutical Markets-1960 to 1999. *Ther. Innov. Regul. Sci.* 2001, 35, 293–317.

[257] Council, N.R., *Toxicity Testing in the 21st Century*, National Academies Press, Washington, D.C. 2007.

CURRICULUM VITAE

Paulien Meert

Address: Gentsesteenweg 78
B-9200 Dendermonde
Place and date of birth: Dendermonde, 06/12/1988
Professional address: Laboratory for Pharmaceutical Biotechnology
Faculty of Pharmaceutical Sciences
Ottergemsesteenweg 460, B-9000 Gent
+32 9 264 80 53
Paulien.Meert@UGent.be

Education

2011 - 2016 **PhD Student**
Laboratory for Pharmaceutical Biotechnology
Ghent University
Dissertation: "Histon code caution: clipping and proper propionylation"

2006 – 2011 **Master in Pharmaceutical sciences**
Ghent University

2006 - 2009	Bachelor Pharmaceutical Sciences	Great distinction
2009 - 2011	Master of Pharmaceutical Care	Great distinction

Dissertation: "Histone cleavage as a posttranslational modification during stem cell differentiation" – Ghent University, Language: Dutch

2000 - 2006 **Science-Mathematics**
Sint-Vincentiusinstituut
Dendermonde

Research fellowships

2011 - 2013: Research fellowship of the agency for innovation by science and technology (IWT) – Flanders (Belgium) (1st mandate)
2013 - 2015: Research fellowship of the agency for innovation by science and technology (IWT) – Flanders (Belgium) (2nd mandate)

Scientific curriculum

Publications

- Pitfalls in histone propionylation during bottom-up mass spectrometry analysis.
  Meert P., Govaert E., Scheerlinck E., Dhaenens M., Deforce D. *Proteomics. 2015 Sep;15(17):2966-71.*
  (*Proteomics* IF 2014: 3.807 – JCR rank 17/79 (Q1) category "Biochemical Research methods")

- Phospho-iTRAQ: assessing isobaric labels for the large-scale study of phosphopeptide stoichiometry.
  Glibert P., Meert P., Van Steendam K., Van Nieuwerburgh F., De Coninck D., Martens L., Dhaenens M., Deforce D. *J Proteome Res. 2015 Feb 6;14(2):839-49.*
  (*J Proteome Res.* IF 2014: 4.245 – JCR rank 14/79 (Q1) category "Biochemical Research methods")

- Histone proteolysis: A proposal for categorization into 'Clipping' and 'Degradation'.
  Dhaenens M., Glibert P., Meert P., Vossaert L. & Deforce D. *Bioessays. 2015 Jan;37(1):70-9*
  (*BioEssays* IF 2014: 4.730 – JCR rank 11/85 (Q1) category "Biology")

- Detailed method description for non-invasive monitoring of differentiation status of human embryonic stem cells.
  Scheerlinck E., Van Steendam K., Vandewoestyne M., Lepez T., Gobin V., Meert P., Vossaert L., Van Nieuwerburgh F., Van Soom A., Peelman L., Heindryckx B., De Sutter P., Dhaenens M., Deforce D. *Analytical Biochemistry* 2014, 461: 60-66.
  (*Analytical Biochemistry* IF 2014: 2.219 – JCR rank 32/76 (Q2) category "Analytical Chemistry")

- Identification of histone H3 clipping activity in human embryonic stem cells.
  Meert P.*, Vossaert L.*, Scheerlinck E., Glibert P., Van Roy N., Heindryckx B., De Sutter P., Dhaenens M., Deforce D. *Stem Cell Research* 2014, 13: 123-134.
  (*Stem Cell Research* IF 2014: 3.693 – JCR rank 31/163 (Q1) category "Biotechnology & Applied Microbiology")

Manuscripts in submission/ preparation

- Preparing histones for label-free quantitative mass spectrometry: a comparison of extraction protocols.
  Govaert E., Van Steendam K., Scheerlinck E., Vossaert L., Meert P., Stella M., Willems S., De Clerck L., Dhaenens M., Deforce D. *Submitted to Journal of Proteome Research*

- The development of a fully defined SILAC culture medium with minimal arginine conversion in human embryonic stem cells.
  Scheerlinck E., Van Steendam K., Govaert E., Vossaert L., Meert P., Van Nieuwerburgh F., Van Soom A., Peelman L., De Sutter P., Heindryckx B., Dhaenens M., Deforce D. *Submitted to Journal of Proteome Research*

- Efficient and specific chemical derivatization of histones by means of propionic anhydride.
  Meert P., Dierickx S., Govaert E., De Clerck L., Willems S., Dhaenens M., Deforce D. *Manuscript under revision in Proteomics*.

- Introducing combinatorial PTM modelling in an improved histone analysis workflow.
  Willems S., Dhaenens M., De Clerck L., Govaert E., Meert P., Van Neste C., Van Nieuwerburgh F., Deforce D. *Manuscript in preparation*.

Poster presentations on national and international conferences

- Evaluation of a Strategy to Identify Protein Phosphorylation in an unbiased way
  Glibert P., Dhaenens M., Meert P., Colaert N., Martens L., Deforce D. *"Belgian Proteomics Association" Conference*, Ghent, Belgium, November 29-30, 2012

- Validation of a non-destructive method to examine human embryonic stem cell (hESC) (non)-differentiation with a OCT4-eGFP Knock In hESC line.
  Scheerlinck E., Dhaenens M., Vandewoestyne M., Van Steendam K., Lepez T., Gobin V., Meert P., Vossaert L., Van Nieuwerburgh F., Van Soom A., Peelman L., Heindryckx B., De Sutter P., Deforce D. *"Knowledge for Growth" Conference*, FlandersBio, Ghent, Belgium, May 30, 2013

- Simultaneously Quantify Protein Expression And Post-Translational Modifications
  Glibert P., Dhaenens M., Meert P., Colaert N., Van Nieuwerburgh F., Martens L., Deforce D. *"Knowledge for Growth" Conference*, FlandersBio, Ghent, Belgium, May 30, 2013

- Histone H3 clipping upon early human embryonic stem cell differentiation
  Vossaert L.*, Meert P.*, Dhaenens M., Deforce D. *"Knowledge for Growth" Conference*, FlandersBio, Ghent, Belgium, May 30, 2013

- Mass Spectrometry top-down characterization of histone H3 on the MS level *(presenting author)*
  Meert P.*, Vossaert L.*, Dhaenens M., Deforce D. *"Knowledge for Growth" Conference*, FlandersBio, Ghent, Belgium, May 30, 2013

- PiTRAQ: a strategy to simultaneously correlate protein expression and phosphorylation stoichiometry between different samples: evaluation on different mass spectrometers

Glibert P., Dhaenens M., Meert P., Colaert N., Van Nieuwerburgh F., Martens L., Deforce D. *"ASMS (American Society for Mass Spectrometry) Conference*, Minneapolis, USA, June 9-13, 2013

- iTRAQ as a method for optimization: enhancing peptide recovery after gel fractionation *(presenting author)*

  Glibert P., Meert P., Van Steendam K., Dhaenens M., Deforce D. *"ProteoMMX3 Strictly quantitative" Conference*, Biochemical Society, Chester, UK, March 25-27, 2014

- Histone H3-protease activity in human embryonic stem cells

  Vossaert L., Meert P., Dhaenens M.? Deforce D. *"Knowledge for Growth" Conference*, FlandersBio, Ghent, Belgium, May 8, 2014

- Histone H3 clipping in hESC in relation to Oct4 expression and culture conditions *(presenting author)*

  Meert P., Vossaert L., Scheerlinck E., Glibert P., Van Roy N., Heindryckx B., De Sutter P., Dhaenens M., Deforce D. "*First meeting of Belgian Society for Stem Cell Research*", Ghent, Belgium, September 12, 2014

- Pimp your (propionylation) protocol: progenesis QI uncouples protocol comparison from the bias of targeted data analysis *(presenting author)*

  Meert P., Scheerlinck E., Govaert E., Dhaenens M, Deforce D. *"Belgian Proteomics Association" Conference*, Brussels, Belgium, December 18-19, 2014

- Evaluation of histone extraction protocols for label-free mass spectrometry

  Govaert E., Van Steendam K., Scheerlinck E., Meert P., Vossaert L., Dhaenens M., Deforce D. *"Belgian Proteomics Association" Conference*, Brussels, Belgium, December 18-19, 2014

Educational experience

- *Supervisor of practical courses Phytochemistry (2nd Bachelor students)*
- *Supervisor of practical courses Pharmaceutical Biotechnology (3rd Bachelor students)*
- *Supervising thesis dissertations*
  - 2011-2012: Leen Feys (1st Master of Pharmaceutical sciences – Ghent University) "Optimalisatie van een kwantitatieve massaspectrometrische techniek in de fosfoproteomics ter bepaling van de activiteit van het US3-kinase"
  - 2011-2012: Stefi Premereur (1st Master of Pharmaceutical sciences – Ghent University) "Humane embryonale stamcellen: Histonklieving binnen stamceldifferentiatie"
  - 2012-2013: Maarten Batens (Master of Science in Industrial Pharmacy – Ghent University)
    "De klieving van histon H3 bij humane embryonale stamcellen"
  - 2012-2013: Ine Coene (1st Master of Pharmaceutical sciences – Ghent University) "Histonklieving H3 tijdens Humane stamceldifferentiatie"
  - 2013-2014: Sarah Dewulf (Master of Science in Industrial Pharmacy – Ghent University) "Onderzoek naar het enzym verantwoordelijk voor histonproteolyse in hESC"
  - 2013-2014: Sandra Van Der Hoofden (1st Master of Pharmaceutical sciences – Ghent

University)
"Analyse van posttranslationele histonmodificaties via chemische derivatisatie en kwantitatieve massaspectrometrie"

- 2014-2015: Sebastiaan Dierickx (1st Master of Pharmaceutical sciences – Ghent University)
  "Increasing identification efficiency and quantitative accuracy in mass spectrometric analysis of posttranslational histone modifications"


UNIVERSITEIT
GENT